On the Trail of a Thought

A Kinenoetic Analysis of Problem-solving

Wendy ROSS

Department of Psychology,

Kingston University, London

December 2020

This thesis is submitted in partial fulfilment of the requirements of Kingston University for the degree of Doctor of Philosophy

Abstract

The research in this thesis describes a microgenetic investigation of thought as it occurs in and through objects and informed by work in distributed cognition and interactivity. The thesis opens with a detailed survey of the arguments in cognitive philosophy around the ontological locus of cognition. I advance the conclusion that many of the open questions will not be solved by empirical methods and suggest a pragmatist approach. Four empirical studies are reported: Three laboratory-based studies which feature traditional problem-solving tasks found often in cognitive psychology and one which examines an artist solving problems which arise over the course of the artistic process. Each of the studies combines quantitative analysis with qualitative analysis of video recorded material to describe thinking in an open cognitive ecosystem. The first study reports performance on a word production task and finds that engagement with external representations is crucial to scaffold performance. The second study uses anagrams to assess the nature of that engagement and concludes a non-agentic model of mere luck is not sufficient. Study three examines performance on an insight problem and suggests that when the problem is not one which is easily scaffolded by material objects, systems form around other types of external scaffold. The final study tracks thought as it unfolds through making of a flower in an artist's atelier. The findings of all the studies support the notion that cognition emerges in the form of material traces and actions on the world.

The thesis introduces and develops two concepts—microserendipity and exaptative action—that offer a new perspective on the nature of problem solving and creativity. These concepts bring in sharp relief environmental chance in creativity when it is enacted; the

methodology employed in the empirical work reported here also permits the identification of events when environmental chance is unnoticed. These phenomena operate outside the conscious observation of the problem solver so they cannot be tracked through traditional methods. The work reported here introduces kinenoetic analysis that trace in micro detail the dynamic transactional coupling between thought and objects that chart the origin of new ideas. The knowledge that the participant generates through the movement of objects mirrors the knowledge gained by the experimenter by these movements. The last chapter introduces kinemorphism as part of a qualitative description of the creative trajectory of an artist working with clay: form is unstable and arises out of action. Such a perspective suggests that what is produced cannot be explained by a reductive process that focuses on only one or the other, but rather must take into account the relationship which arises through action. Creativity from this perspective is transactional and relational. In terms of theoretical contributions, I cast doubt on an agent centric view of interactivity which posits an uncomplicated augmentative relationship between things beyond and within the brain and suggests instead a transactional approach to knowledge acquisition. These lead to novel observations on the role of the experiment in research in situated cognition. Reflections on the pluralistic method of kinenoetic analysis are offered and directions for future research are outlined.

.

Acknowledgements

The work in this thesis suggests that the process of discovery is distributed across people and things and there is no truer reflection than the number of people that I have to thank for their contribution.

First, of course, and most important, my supervisor Professor Frédéric Vallée-Tourangeau who has been a constant source of inspiration and academic generosity. I will forever be grateful. I sincerely hope that this is only the start of our collaborative journey. Thank you for changing my story.

Dr Samantha Copeland has been a constant course of advice and mentorship as well as offering robust critical debate over areas I am still stumbling over. It gives me immense happiness to call her my friend. The work we do together is a constant source of joy to me. I am excited to see what the future will hold for us as the work we do grows closer. We have built strong foundations of which I am very proud. She also gave detailed feedback and guidance on Chapters One, Two and Three.

Dr Vlad Glăveanu has been nothing but responsive and generous with his time and advice. There is nothing lovelier than his supportive and validating messages and yet the intellectual robustness of his work is equally breath-taking. As a mentor and guide, he has been an invaluable influence. He gave close feedback on Chapters Six and Seven but regular, at times daily, conversations have left their mark throughout and his influence is woven through this thesis as a whole.

As well as lending me his time and flower making skills in Chapter Seven, Paul March has pushed me intellectually in directions I never could have imagined. I have been exceedingly lucky to have been able to meet and work with someone who is so immensely talented. It is one of the great pleasures in my life to have witnessed his creative process. His thoughtful comments on early drafts of the chapter, his fearless and principle intellectual positioning and his unfailing emotional support have made this possible. Again, his influence has shaped so many of the thoughts in this thesis.

Dr Margaret Webb has been my inspiration personally and professionally, negotiating with me the pitfalls of early career researchers and guiding me through the literature on insight research. Despite the time zone difference, our reading groups and work together has helped me to appreciate the nuances of research which develops out of a different theoretical background.

Conversations with Dr Selene Arfini and collaborative writing with her has strengthened my convictions that an externalist approach to creative cognition is not only pragmatically the clearest way forward but the only conceptually coherent approach. It was a delight to work with her and I hope to do so many times in the future. She has shown me the benefits of interdisciplinary collaboration in practice.

Robust debates and collaborative writing with Dr Sam Smith and Dr Marcus Dahl (JE Vistic) have broadened my understanding of the balance of the individual and the society in creativity and taken me back into deep time in my thoughts about how we can understand cognition. I am incredibly grateful to them both for the time, good humour, fun and endless unread but shared papers.

The support and friendship from the research community at Kingston University has got me through some of the more troubled times preparing a PhD during a pandemic. Dr Rose Turner held my hand as I registered for the PhD and has blazed the trail showing me how a true academic should be. A special thank you to Ifigeneia Manitsa and Amy Montague for support and laughter, so much laughter. Motorbikes, racks and balls of all sorts will always have a special place in my heart. Thank you to Milani Pathmanathan for all the love. Thank you to Susan Cooper for thoughtful comments on Chapters Two and Five and for thinking through knotty problems of all sorts with me. Her support at the crunch times in all ways was invaluable. Andrea Marin offered thoughtful commentary on the video data for Chapter Five.

I have been lucky enough to enjoy the support in the final few months of the Organisational Psychology department at Birkbeck College. The intellectual curiosity and generosity extended by the staff members there throughout the final few months of thesis writing and the support offered as a I prepared for my viva was invaluable. Special thanks to Dr Susan Kahn, Dr Janet Sheath and Dr Kevin Teoh for their support over submission and viva prep.

I have greatly enjoyed working with the students I have been privileged to teach and none more so than Renata Ruffatto Rech. She offered thoughtful comments on Chapter Two alongside being a source of inspiration throughout. I am so immensely proud of the academic she is growing into. I look forward to seeing her spread her wings.

Anthea Salisbury's stability and patience with my crazy ideas have definitely kept me sane. Our circular walks around Gigg's Hill Green while trying to cope with the lock down, our silly challenges, her kind patience. Even upside down she makes everything feel the right way up again. I owe her a great debt.

James Robertson, without whom I may never have got it out of the locker. His energy pushed me through the last few months of writing a PhD in a pandemic. Thank you for accepting me just as I am.

Frédérique Barret, Nicole Murray, Louise Fleming, Lucy Seeley, Helen Clare, Carys Galer, Cat Mackey, Greg Mina, Rob Bouffe, Will Fawcett, Darren Smilie, Lewis Watson, Michael Spencer and many others. The world is a better place because you are in it.

Finally, my dearest family. My parents have always been there with unfailing support and pride. The patience of my children, Sam, Sophie and Will, not only with my protracted absences but more so with my crazy ideas. I am so proud of the gorgeous people they are turning into.

And Max, the old dreammaker. I am thrilled to have him by my side during this whole adventure. Thank you, my love, for all your support and oft-times bemused patience. I can still see the same rainbow’s end.

Online Supplementary Information

This thesis has an Open Science Framework webpage associated with it which provides a repository for reports of pilot studies as well as the data files for each of the experiments. The OSF webpage can be found here: http://tiny.cc/RossPhD

Chapter Four

Appendix A	Pilot Task Report
Appendix B	Main study quantitative data
Appendix C	Frame by frame coding of cognitive cases

Chapter Five

Appendix D	Pilot Task Report
Appendix E	Main study quantitative data
Appendix F	Examples of a qualitative memo from the nested data (P10)

Chapter Six

Appendix G	Quantitative data
Appendix H	Examples of two qualitative memos

Chapter Seven

Appendix I	Transcript of dyadic interviews
Appendix J	Transcript of *If This is Not Art*
Appendix K	Video of the flower making

Associated Publications

Parts of this thesis have been reproduced in the following publications. Copies of these publications are also available in the Online Supplementary Information: http://tiny.cc/RossPhD.

Ross, W., & Vallée-Tourangeau, F. (2019) Unknitting the meshwork: Interactivity, serendipity and individual differences in a word production task. In A. Goel, C. Seifert & C. Freska (Eds.), *Proceedings of the Forty-First Annual Conference of the Cognitive Science Society* (pp. 2674- 2680). Cognitive Science Society.

Ross, W. (2020). Serendipity. In V. P. Glăveanu (Ed.), *The Palgrave encylopedia of the possible*. Palgrave MacMillan.

Ross, W., & Vallée-Tourangeau, F. (2020). Catch that word: Interactivity, serendipity and verbal fluency in a word production task. *Psychological Research*. https://doi.org/10.1007/s00426-019-01279-y

Ross, W., & Vallée-Tourangeau, F. (2020). Microserendipity in the creative process. *Journal of Creative Behavior*. https://doi.org/10.1002/jocb.478

Ross, W., & Vallée-Tourangeau, F. (in revision). Accident and agency: A mixed methods study examining. *Thinking & Reasoning*.

Ross, W., & Vallée-Tourangeau, F. (under review). Rewilding cognition. *Journal of Trial and Error.*

Ross, W., & Vallée-Tourangeau, F. (under review). Kinenoetic analysis: Unveiling the material traces of insight. *Methods in Psychology*.

Ross, W., & Arfini, S. (Forthcoming). Serendipity and creative cognition. In L. J. Ball & F. Vallée-Tourangeau (Eds.), *Routledge handbook of creative cognition*.

Contents

List of Figures

List of Tables

Notes on Software Used in the Thesis

Quantitative data were managed in Microsoft Excel (365)

All statistical analyses for Chapters Four and Five were were conducted in IBM SPSS v26.

For Chapter Six they were conducted in JASP (v.0.14; JASP Team, 2020)

All power analyses completed in GPower (v.3.1.9.4; Faul et al., 2007).

Video data for Chapters Five was analysed in ELAN (v.5.4; Max Planck Institute for Psycholinguistics, 2018) and for Chapter Six ELAN (v.5.5; Max Planck Institute for Psycholinguistics, 2019)

Eye tracking data for Chapter Seven was collected using Tobii 2 Glasses. The video was exported to ELAN (v.5.5; Max Planck Institute for Psycholinguistics, 2019)

General Overview

This thesis reports the findings of four empirical studies which take different approaches to problem-solving. Chapter One surveys the philosophical literature with regards to cognition and comes to the conclusion that a pragmatist approach which treats cognition as distributed across people and things is most suitable, that is one which maintains a focus on what people actually *do* in problem-solving. This position forms the theoretical basis for the rest of the thesis. Chapter Two takes a closer look at theories of problem-solving and discusses the problems with using as a tool the very thing you are aiming to measure, as well as offering some suggestions to resolve the potential regress. Chapter Three moves to methodology and assesses the dominant methods in cognitive investigations of problem-solving before suggesting a mixed methods approach based on quantitative analysis followed by behavioural and critical case analysis. The empirical bulk of the thesis is reported in Chapters Four to Seven. Chapter Four reports the results from a study which required participants to create as many words as they could out of a set of letters. Chapter Five follows up on this by asking participants to solve one solution anagrams. Chapter Six reports the results of an experiment which did not elicit the desired manipulation and uses the video data gathered to discuss reasons for this and discusses the difficulty with experimental manipulations of open systems. Chapter Seven moves away from the laboratory and reports a different form of creative problem solving, namely that arising from actions on and with clay in an artist's atelier. Chapter Eight discusses the implications of the methodological innovations and the explanatory mechanisms that have been uncovered before suggesting how the findings from this thesis can direct a new research programme.

Chapter One : Approaches to Cognition

This thesis deals with how agents solve problems as a lens to make broader comments on approaches to measuring cognition and especially higher order and creative cognition[1]. The overarching research aim is to uncover whether there are explanatory factors that are not accounted for by our current models of problem-solving and creativity. To do this I will explore new methods that focus on how knowledge can be generated by, and mapped through, the movement of objects in the world. This represents a departure from traditional approaches to cognition. Such approaches view mind and world as separate and suggest that the workings of the mind can be uncovered through models which rely on a rational and computational approach. The overall argument of this thesis is that these models do not represent the idiosyncratic and contingent nature of behaviour in open cognitive ecosystems.

Problem-solving is so entangled with the philosophy and psychology of cognition that at a recent talk, Wheeler (2020) exclaimed: "That's what cognition is, it's problem-solving and that's really a very deep assumption in most cognitive psychology" (25.06) and Dixon et al. (2014, p. 160 emphasis added) suggest that "problem-solving provides us with a microcosm of the *key issue* for psychology: 'How does an organism successfully engage in goal-directed action?'". So, research on problem-solving is entangled with and recursively informs research on cognition and both are constrained by, and feed into,

---

[1] There is no clear definition of what constitutes higher order cognition. Zagaria et al., (2020, p. 527) present an analysis of psychology text books and conclude: "Cognition is so variously formulated that it is difficult to summarize its definitions. Broadly, cognition seems to be an "umbrella-term" under which all the activities traditionally considered to be "cold cognition" are grouped: e.g. information processing, thinking, reasoning, problem-solving, understanding, knowing." In practice, higher order cognition is seen as cognition which requires effortful processing. While the presented here aims to question the boundary between higher order and lower order cognition, I follow Kirchhoff (2018, p. 6) who suggests that examples of higher order processes would be : "cognitive control, memory, decision-making, social cognition, mathematical reasoning, and language understanding." In short, those things which are considered processes of the intellect.

theoretical traditions in cognitive science. Save Chapter Seven, the tasks employed in the research presented here are simple problem-solving exercises typically employed by cognitive psychologist interested in the science of problem-solving. However, with simplicity comes clarity: the procedures developed offer a fruitful platform from which to examine how coupled agent-environment systems gain new knowledge about a situation. through action whether goal directed or not.

This thesis is undergirded by the premise that the classical approach to cognition is not sufficient to fully understand the activities of a situated problem solver and takes the stronger view that thought is constructed and constituted by actions over objects and finds these object-thoughts can be traced. The thesis will invite a re-examination of some of the more foundational aspects of experimental research in cognition and will introduce explanatory factors which cannot be accounted for in the classic models. To situate this position more clearly, I start by exploring the different approaches to cognition that can be found in the literature. I will begin by outlining the cognitivist perspective which provides the foundation for cognitive approaches to psychology before outlining four major alternative approaches: the theory of the extended mind, embodied cognition, enactivism and ecological approaches. Finally, I shall outline the theoretical basis of, and the research in, interactivity which forms the theoretical approach of this thesis.

A Cognitivist Perspective

Broadly speaking cognitivism puts the locus of cognition solely inside the person and proposes that cognition consists of internal computations over mental representations[2] (Rowlands, 2010; Wilson, 2001). It has two main tenets. First, that thinking supervenes on internal, physical states. In other words, that understanding internal states is both necessary and sufficient for understanding thinking and that changes in one necessarily cause changes in the other. Second, that thinking (that is the "thought-full" processes that make up cognition) is best viewed as information processing modelled on the way a computer processes information. In short, on this view cognition consists of the processing of internal representations according to rules which are, in theory, discoverable (Clark, 2014; Fodor & Pylyshyn, 1988; Lycan & Prinz, 2008). Human cognition, therefore, from the cognitivist position, is the computation of information; it proceeds in the same way as a computer and adheres to rational and discoverable universal principles. This is the fundamental starting position of much of mainstream cognitive science (Noë, 2006) and leads to the cornerstone aim of cognitive science: the discovery of the rules which govern internal computational representationalism (Núñez et al., 2019).

---

2 Mental representations are complex and well debated and a full account of these debates is beyond the scope of this thesis. Fodor memorably describes the position thus:

> It rained for weeks and we were all so tired of ontology, but there didn't seem to be much else to do. Some of the children started to sulk and pull the cat's tail. It was going to be an awful afternoon until Uncle Wilifred thought of Mental Representations (which was a game that we hadn't played for years) and everybody got very excited and we jumped up and down and waved our hands and all talked at once and had a perfectly lovely romp. But Auntie said that she couldn't stand the noise and there would be tears before bedtime if we didn't please calm down." (Fodor, 1985, p. 76)

In short, representations are the internally generated reflection of reality. The underlying basis of a Representational Theory of Mind is that given all stimuli are necessarily converted to representations, the focus of research should be on the nature and workings of those representations. This leads to the approach of methodological solipsism which I discuss in more depth later in the chapter. The relationship between mental representations and cognitive processes and the external world is complex and underlies much of the thesis. The nature and philosophical status of the representations are less important.

It is important to note, as Wallace (2007) does, that this way of looking at the mind was not a discovery in the same way that other scientific discoveries such as those from astronomy were, it is "not a self-evident fact" (Dixon et al., 2014, p. 71). It is a heuristic employed to better support understanding of what cannot be seen or otherwise accessed. We cannot see thought nor thinking. While it is a foundational assumption of cognitivism it is not a *necessary* assumption. Cognition, at its root, is concerned with the generation of knowledge[3], the location or manner of that knowledge generation is not predetermined other than by the allegiances[4] of the researcher. Indeed, it is widely noted that the computer is only the most recent of a series of culturally contingent metaphors to describe the mind - others include a blank tablet ready for impressions or a telephone switchboard (Clark, 2014; Hutchins, 1995; Thagard, 2005). The *tools to theories* heuristic outlined by Gigerenzer and Goldstein (1996) demonstrates how tools are not only designed to test scientific theories, but they are also instrumental in the generation of the theories, in this case the conceptual change about the nature of human cognition was informed by, and created in tandem with, the rise in computing and computer programming. The cognitive revolution[5] of the 1960s is inextricably entwined with a computing revolution and the search for human intelligence entangled bidirectionally with the search for artificial intelligence. Epistemology and ontology are entwined.

---

3 While Merriam-Webster defines cognition as "cognitive mental processes", the word comes from the Latin root of cognoscere meaning "to know or to investigate" (Hoad, 1993, p. 83).
4 These allegiances can vary in nature from unthinking disciplinary traditions to staunchly held theoretical positions. As I shall argue in Chapter Eight, these allegiances and positioning underlie both qualitative and quantitative approaches even if it is only the former which recognise this.
5 Note that this only a revolution in American psychology; psychological research in Europe had never been quite so much in sway to the behaviourist promise (Mandler, 2011) and during that time Bartlett published *Remembering* (1932) while Kohler introduced the idea of insightful problem-solving in *The mentality of apes* (1925). In the 40s and 50s highly cognitive theories of attention were also being developed by those such as Broadbent (1958) and Treisman (1960) who were inspired by the work of Craik (1943) on mental models.

The words we use to understand human cognitive processes such as memory and thinking are indistinguishable from those used to understand a computer programme such as input, encoding, retrieval, processing or output. When Plato famously invites us to compare the mind to an aviary[6], he is not suggesting that the brain literally functions as a contained group of birds. However, when it comes to the metaphor of a computational model, common talk of output and input are taken seriously by researchers (Hurley, 2002)[7]. Most cognitive scientists adhere to the view of the brain as a computer, that is that cognition can be explained in terms of computations, a position which may be controversial but which is nonetheless resilient (Dixon et al., 2014; Piccinini, 2012).[8] As

---

[6]In Theaetetus (http://classics.mit.edu/Plato/theatu.html)

[7] Since Baddeley and Hitch's (1974) model of working memory, there is a certain type of undergraduate, indeed postgraduate at times, student who will seek to find the physical location of working memory. First, this points to an interesting inability to fully abstract which is important in how we understand scientific models. If there is a long-term store, well, where is that store? Can we access it? Can we hack it in some way, extend the RAM? The functional boxology common in cognitive neuropsychology modelling that assigns different areas of the brain to different functions—which inevitably veils the distributed nature of neural networks and resists neuroplasticity—reinforce this conceptual misunderstanding (Anderson, 2014). More seasoned researchers and thinkers will scoff knowing that working memory is not something that occupies a physical space beyond the numbers on an excel spreadsheet (or indeed an R matrix) it is only a skill which can be evidenced. It is a portmanteau term for a series of cognitive processes which influence performance on a working memory test and on other tasks which require working memory. There is an obvious problem here with the limits of language; when we speak of working memory 'capacity' or the longer memory 'store' we are invited to consider these things as occupying a physical space. Our relations with the abstract are mediated through concrete embodied metaphors which enable less concrete thinking (Lakoff & Johnson, 2010). Therefore, the shadow of the physical is ever present. It is for this reason that we should be wary of some of the arguments from the philosophy of the extended mind which we will address in the next section, these examples invite us to draw parity between a metaphor based on a tangible object and a concrete and measurable thing. For example, Clark and Chalmers's (1998) thought experiment asks us to consider whether a notebook owned by Otto plays the same role as memory in his counterpart Olga (see p. 15). The argument that a notebook should be seen as part of his cognitive system it also implicitly invites us to view Olga's memory as if it were also notebook. The very real risk of bringing concrete objects into the cognitive system is that it invites us to drop the idea of internal representations as a useful metaphor and suggests that it is as real and measurable as external representations. Part of the argument for extending the cognitive system is that the external world is not the same but often different and can be either facilitative or not.

[8] Although some of the explanation of that resilience has to be attributed to the rather wide definition and list of unresolved complexities: "…much work remains to be done to characterise the specific computations on which cognition depends: whether we ought to be committed to weak or to strong associationism: whether—and to what extent—the satisfactory explanation of cognition requires classical computational mechanism as opposed to non—classical digital computation, as some have maintain. It may turn out that one computation theory is right about all of cognition or it may be that different cognitive capacities are explained by different kinds of computation. To address these questions in the long run, the only effective way is to study nervous systems at all its levels of organization and find out how they exhibit cognitive capacities." (Piccinini, 2012, p.243)

Fodor (1980, p. 65) writes, "[…] we may thus construe mental operations as pretty directly analogous to those of a Turing machine".

The metaphors used by a scientific research programme provide the means by which scientists structure those aforementioned research programmes (Fernandez-Duque & Johnson, 1999; Koriat & Goldsmith, 1996). Most important, this computational view of cognition allows cognition to be conceived as a series of self-contained modules that can be parsed in a sensible and meaningful way *in isolation* (Gigerenzer & Goldstein, 1996; Karmiloff-Smith, 1992; Robbins, 2017). Fodor (1980) argues that the only sensible approach to understanding cognition requires this "methodological solipsism" (see also Wilson, 2001). In other words, that the world of the brain is enough to understand cognition because the external world is transformed into representations within the brain with no necessary external referent. This is a strong commitment to uncovering mental states without reference to their representational content. He does not deny an external reality, but he does deny the logic of a methodological commitment to an investigation of these relations and what they represent: "[…] there's no practical hope of making a science out of this relation. And of course, for methodology, practical hope is everything." (Fodor, 1980, p. 71).

By isolating cognitive processes in this way, things 'outside the brain' could be addressed once the more fundamental matters of internal cognition had been established (Hutchins, 1995). So, it is not that the external world was, or continues to be seen as, irrelevant—indeed it is trivially true to suggest that thinkers are embedded in a context—but rather it is not necessary to understanding the rule-governed structure of human thought and, perhaps more important, will act as a complication to the discovery of these universal laws. Indeed, this is the approach adopted by Simon (1965) who introduces us to the idiosyncratic behaviour of an ant crossing the beach and demonstrates that this behaviour is mainly determined by the surrounding environment:

> An ant, viewed as a behaving system, is quite simple. The apparent complexity of its behavior over time is largely a reflection of the complexity of the environment in which it finds itself. (p. 52)

Simon uses this analogy to suggest that behaviour, human as well as ant, is not a reflection of thinking as effortful and intellectual, but rather it is a reaction to chaotic environmental factors. Counterintuitively, Simon goes on to reject the idea that such an observation therefore requires a systemic and interactive approach to measuring behaviour. He argues that the proper study of cognition is the nomological pursuit of behaviours which are not dependent on environmental complexity to be expressed and are stable across different situations. Indeed, the theory he advances has two caveats ("hedged bets"): First, that this does not apply to an embodied thinker and second, that "information packed memory" should be considered part of the system rather than the organism. This is a reductionist agenda (disembodied and ahistorical) designed to uncover basic universal laws which can then be scaled up to explain more complex phenomena.

For Simon, the environmental complexity makes it harder for us to deduce what is 'cognitive' and what is environmental. Therefore, the proper task of psychology is to only focus on those things which are not attributable to the environment or socio-cultural forces. Experimental psychology should focus on elucidating the parameters of "broad commonalities in organizations of the human information processing system as it engages in different tasks" (Simon, 1965, p. 81). This, of course, assumes that an understanding of the internal, universal properties of cognition is useful without reference to the complex context in elucidating processes which undergird thinking. However, on this view, for the scientific domain of cognitive psychology, the interests and goals of that domain dictate that we look only inwards. While this seems intuitive and respects folk ideas inspired by introspection, it is important to keep in mind that these goals and interests are not natural kinds but rather are conjured through socio-historically situated discourse (Mandler, 2011).

The emphasis on this socio-cultural contingency of the internalist conception of cognition is important because the exhortation to ignore the complexity of the environment was not merely a theoretical aside, it had profound implications on how thinking was and continues to be researched. The act of cognition is measured in its physical absence by experiments. The experiments are predicated on a linear, normative model of cognition and therefore attempt to create a space which could only be logically filled by a certain cognitive procedure. These experiments are administered in a controlled environment which is assumed to be sterile, free of any environmental influence which might add complexity and undermine the generalisability of the results. That this is not actively considered a theoretical decision does not belie its theoretical underpinnings (Noë, 2010).

For example, there are three main models of arithmetic cognition (that is, models to explain how people across a broad developmental span do arithmetic). Two—the Abstract Code Model(McCloskey, 1992; McCloskey et al., 1985) and the Triple Code Model (Dehaene, 1992) are entirely internal modular theories and view mathematical cognition as converging on an inherently abstract numerical code. The third posits an interaction between the representation of the problem and internal codes—the Encoding Complex Model proposed by Campbell and Clark (1988). These three models all share the fundamental assumption that the research area of interest is the manner in which the external stimuli (numbers) become internal representations and the computational processing that ensues after the internal representations are formed. This processing is assumed to follow normative, rational algorithms.

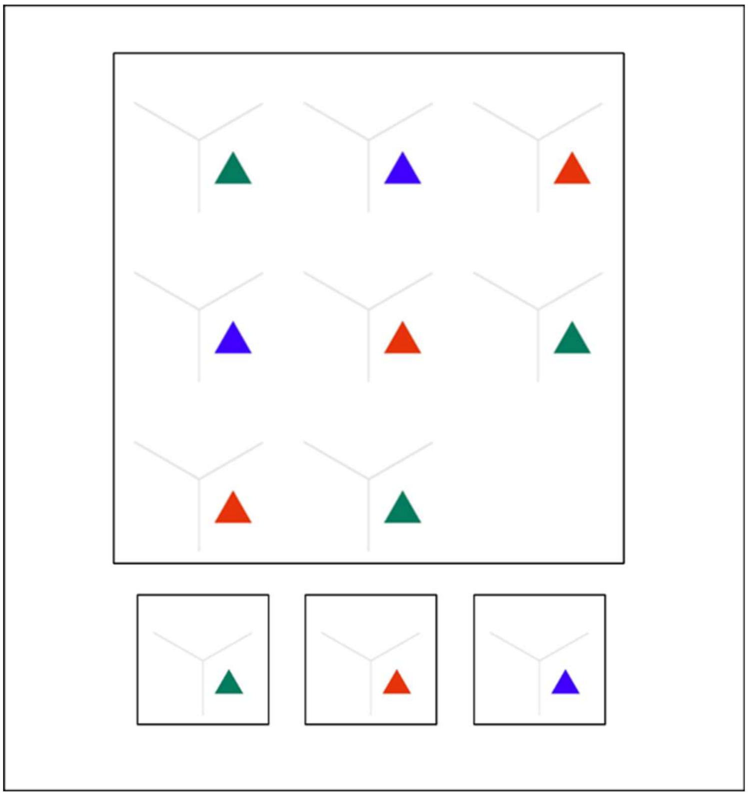
These internalist models restrict our understanding of mathematical cognition. On this view mathematics requires a thinking agent to convert the symbols (the input) into an internal code or representation and perform operations (processing) on them before reconverting this code and stating the answer (output) following linear algorithms. The only question is how that conversion occurs. Evidence that mathematical cognition is structured differently and non-normatively when participants are observed out of the school room (Nunes et al., 1993) or the laboratory (Lave, 1988) or that mathematical cognition is supported and changed by gesture (Goldin-Meadow et al., 2009) or manipulable artefacts (Martin & Schwartz, 2009) is not taken as relevant: models of arithmetic cognition have remained quite immune to these findings.

Limitations of the Classic Approach

The Relationship Between Models and Behaviour. When the underlying cognitive processes posited do not map onto real world behaviours, researchers face a dilemma: Either abandon the nomological computational models or marginalise the relevance of findings that are not captured by these models. Take, for example, intelligence as measured by performance on a common, nonverbal, matrix reasoning task. These tasks measure abstract reasoning as a fundamental cognitive skill which predicts academic performance and, even, mental health (Chierchia et al., 2019). Participants are presented with a grid of pictures with one missing and are asked to select one of the presented options to fit in the gap (see for example Figure 1.1)

Figure 1.1

A Typical Matrix Reasoning Task (Adapted From Chierchia et al., 2019). The Goal is to Select the Missing Shape from the Selection below the Grid.



This task is traditionally measured as a static activity with the number of correct answers and, sometimes, latency taken as dependent measures; these two measures are considered as good predictors of problem-solving and reasoning capabilities (Laurence et al., 2018). Participants can solve the problem either through deducing an initial rule or by using trial and error. However, when participants are invited to move the tiles in a dynamic interface then the predictive power of the test increases (Bocanegra et al., 2019). Furthermore, the patterns of movement that Bocanegra and colleagues mapped demonstrated support for previous predictions (Carpenter et al., 1990) that optimal solutions would require a recursive strategy (one which involves rapidly adjusting goal setting). This optimal strategy is far easier to enact in an interactive environment (see also

Chapter Four for the relationship between environment and strategy which reports supporting findings). It remains untested but it is not implausible that there is a different optimal strategy in the interactive and the non-interactive version and the skills required to enact the optimal strategy in the interactive version are more predictive of overall academic achievement. In this instance, a theoretical commitment to cognition as abstracted leads to a shallow reflection of what intelligence looks like as it unfolds in the context of academic achievement. The standard tests' predictive power is hampered by a failure to consider interactivity.

The Mereological Fallacy

Beyond the possibility that the models cannot account for any as yet unobserved behaviours, by considering cognition as formed of discrete modules which can be logically isolated from environmental complexity, there is the not inconsiderable chance that cognitive scientists risk committing the mereological fallacy (Bennett & Hacker, 2003). Mereology is the study of part/whole relations and this fallacy refers to ascribing to a part what is properly only ascribable to a whole.. The underlying ontological assumption of mereological composition is that the whole is the sum of its parts and nothing more and that through the process of composition or decomposition the parts remain the same. Rocca and Anjum (2020) compare it to building a model with Lego bricks: There is a logic to an analysis of the bricks as separate items in order to understand the whole model because the bricks remain unchanged throughout the process. This same logic underlies the isolation and examination of individual cognitive processes. As I have demonstrated, an internalist perspective relies on the importance of internal states so that changes in the outside environment are mirrored by changes in internal dispositions and so this environment can be ignored. It is this position and the concomitant mereological assumptions which dictate the level of analysis of the research into cognitive processes. If, however, the Lego bricks were malleable and changed during the building process, assessing them at the level of the brick would be meaningless because what would be of interest would be the shape and composition during building.

Alternatives to Cognitivist Approaches

The extent to which cognition can be studied in isolation from the world is not a new consideration and getting the balance right has marked much of modern psychology. For example in 1890, James (p.13) writes

> On the whole, few recent formulas have done more service of a rough sort in psychology than the Spencerian one that the essence of mental life and of bodily life are one, namely, 'the adjustment of inner to outer relations.' Such a formula is vagueness incarnate; but because it takes into account the fact that minds inhabit environments which act on them and on which they in turn react; because, in short, it takes mind in the midst of all its concrete relations, it is immensely more fertile than the old-fashioned 'rational psychology' which treated the soul as a detached existent, sufficient unto itself, and assumed to consider only its nature and its properties.

As we have seen, while it dominates thinking in cognitive psychology, the computational representationalist approach is not a necessary method of studying cognition and various approaches have been proposed which aim to capture the complexity of thinking when coupled with environmental influences. The current challenges to a cognitivist attitude towards the mind are often brought under the catch-all term of 4E cognition where the E represents embedded, embodied, enacted or extended cognition. These positions are grouped together mainly by virtue of their opposition to the dominant framework of cognitivism. While they hold some similar positions, there are other conceptual differences and indeed, metaphysical inconsistencies. The grouping together is perhaps unhelpful and certainly erases much difference (Newen et al., 2018; Rupert, 2004). However, it is still the case that research done in this tradition tends to have to justify its ontology whereas the work which assumes an internalist theoretical base proceeds without explicitly articulating or defending the implicit ontological commitments.

So, despite the initial comfort in a simple external/internal divide, it is also not easy to simply assign these different positions to clear groups; there is a necessary spillage from one to the other reflecting a spectrum of views on important conceptual arguments rather than clear allegiances. What follows is a brief tour of the major positions in this area to better situate the theoretical commitments underlying the work presented in this thesis.[9]

Agent Based and Computational Approaches: The Mind Extended

Early work from McClelland and Rumelhart (1987) first drew attention to the importance of manipulating the external world to support cognitive processes. The authors argue that when it comes to problem-solving, rather than processes reflecting computations over internal representations, this skill is better viewed as underpinned by three human characteristics: pattern spotting, modelling the world and manipulating environments. These abilities allow us to literally perceive the answers to problems outside of the brain. This final skill (manipulating the environment) is, according to McClelland and Rumelhart, is “perhaps the crucial skill which allows us to think logically, do mathematics and science and in general build a culture.” (p. 44). In their model representations are still processed, but those representations can also consist of an actual physical representation generated by manipulation of the environment. This manipulation therefore allows complex problems to

[9] Material Engagement Theory (Malafouris, 2013) is an additional key theoretical position which informs the research in Chapter Seven and a description of that programme appears there.

be divided into a series of simpler ones. Take for example, a long multiplication problem such as 3372 x 294. This is beyond the capabilities of most people's internal mental resources but given pen and paper and the culturally transmitted knowledge of long multiplication, a primary school child could complete the sum.

The hypothesis of extended cognition was perhaps most famously articulated by Clark and Chalmers (1998) although it draws on evidence from Kirsh and Maglio (1994). The thesis outlined in the 1998 essay is both metaphysically radical and empirically conservative. Clark and Chalmers introduce us to two thought experiments. The first involves four Tetris[10] players. One uses only mental processes to decide how to rotate the zoid most efficiently; she is contrasted with one who uses the gameplay to quickly rotate the zoid in front of her. It is obvious from this that one is clearly mental rotation, and one is clearly external rotation. Two further players are then introduced to muddy the philosophical waters. First, a human from the future who has a retinal implant which displays the zoids in the same way as on a computer screen and which rotates the zoids on the basis of a neural command and second, a Martian who has the same biotechnological machinery occurring naturally. The existence of these two 'in between' cases cast doubt on the binary opposition which classifies that which happens *inside* the brain as mental and that which takes place *outside* the brain as non-mental. From this they offer the Parity Principle which in essence is the spread of epistemic credit beyond the biological boundaries of the human brain:

> If, as we confront some task, a part of the world functions as a process, which, were it done in the head, we would have no hesitation in recognising as part of the

[10] Tetris is a computer game in which the players must quickly decide where to efficiently place falling blocks (zoids) of various shapes on top of the already placed blocks to maximise the space. The blocks can be rotated to fit. The game initially starts slowly but quickly advances in speed so the zoids fall faster thus it is easy to learn but remains challenging even with increasing skill levels.

> cognitive process then that part of the world is (we claim) part of the cognitive process. Cognitive processes ain't (all) in the head. (Clark & Chalmers, 1998, p.8)

This is then extended with the example of Otto and Inga. Both wish to go to an art exhibition at the Museum of Modern Art in New York. Inga has a normally functioning memory system, which she consults before going to the museum on 53rd Street, Otto has Alzheimer's but carries a notebook on which he relies absolutely and which tells him the location of the Museum of Modern Art. The authors argue that under the right conditions (tightly specified and argued over at length), the condition of dispositional believing could be extended beyond the brain boundary.

When it comes to empirical evidence, all that is required to support the argument from Clark and Chalmers that cognition[11] functionally extends is evidence of recruiting the environment as a cognitive offload (Risko & Gilbert, 2016). So, in the case of mathematical cognition, for example, a cognitively extended system is one that makes use of a range of external resources from pen and paper (Cary & Carlson, 1999) to a calculator (Ruthven & Chaplin, 1997) to expand the mental workspace. However, according to the argument made here mathematical cognition can also proceed offline[12] on the basis of internal representations even if the overall efficiency of the system is reduced (Wheeler, 2004). This is an agent centric version of the situated mind, the internal and external boundary remain intact even while the cognitive system expands (Clark, 1997).

Embedded or Extended Cognition?

There have been many arguments and counter arguments around the extended mind

[11] Clark and Chalmers (1998) make the argument for the extension of cognitive processes and beliefs. This latter requires more empirical evidence and has become largely disregarded although it is central to the argument of the essay.

[12] By "offline", I mean in the absence of external, scaffolding stimuli.

and Clark and Chalmers have shifted position and issued clarifications many times over the more than 20 years since its publication (see Colombo et al., 2019 and Menary, 2010 for useful collections setting out support and criticism). However, there is no doubt that it has founded a rich research area and had acted as a canonical text for theories of extended mind. One of the closest positions empirically, and yet furthest metaphysically, is the argument for embedded cognition. Proponents of the embedded mind thesis broadly agree with the hypothesis of the extended mind with one simple caveat: While the external world can be an important *causal* factor in cognition, it is not *constitutive* of cognition. This means that that the main theorists in this camp (particularly Rupert, 2004, 2009) agree that the explanation for most intelligent behaviour will need to include the environment in its explanation but that this doesn't make the environment a cognitive agent. There is no mark of the cognitive which can attach to non-organic things. Thus, proponents of embedded cognition will reject the metaphysical claim of the extended mind thesis but do accept a causal dependence on the environment. As I shall argue later, because the extended and embedded minds both adhere to a agent-centered and computational model of cognition, there is little difference in empirical predictions even though the claims on the ontological locus of cognition are irreconcilable and diametrically opposed.

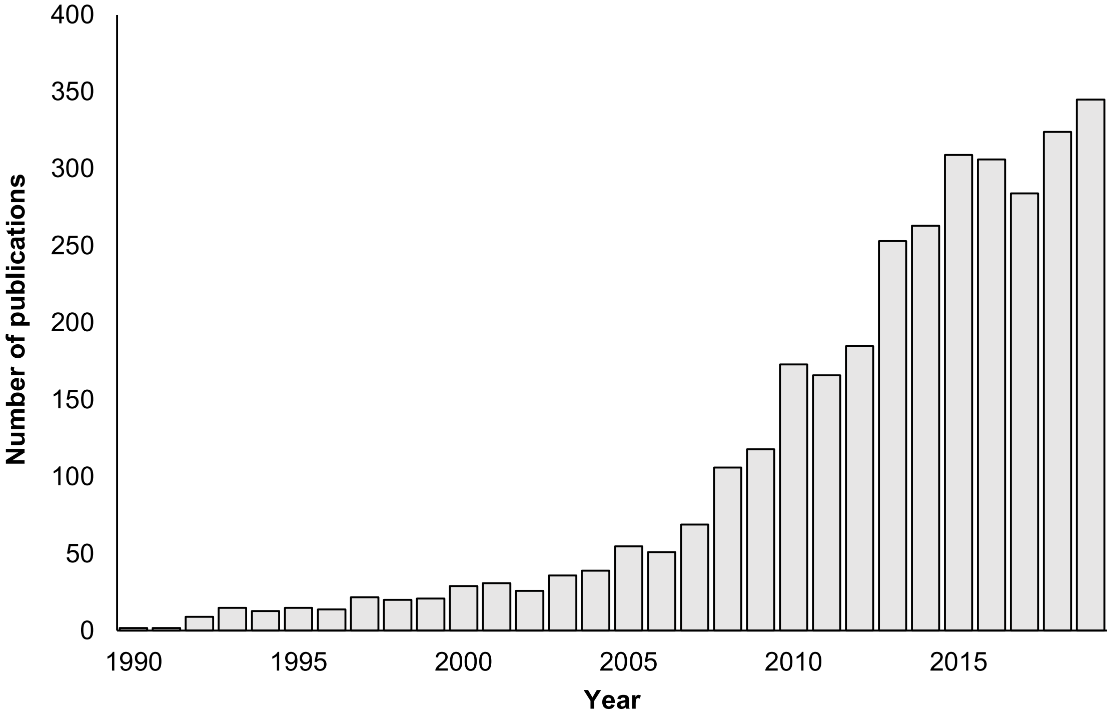
Embodied Approaches: The Mind Decentred

Embodied cognition is perhaps the broadest church in the 4E community and also one of its more prolific. An embodied stance has gained a lot of traction within the cognitive science community: There has been exponential growth of journal articles and published proceedings since 1990 which shows little sign of abating. This rise is illustrated by Figure 1.2 (inspired by a similar analysis in Mahon, 2015). In all cases of embodied cognition, cognition is distributed across the body. Extended and embedded mind theories maintain the brain as the centre of cognition, while the work of embodied cognitive scientists shift the centre to the bodily processes which do not feature in other cognitive models. For example, Butterworth (1999) suggests that the use of fingers in learning to count is important for developing a sense of numerosity and this is supported by research in manumerical cognition, or finger counting. Despite many educationalists' disapproval, the use of fingers to learn to count is particularly important both for children understanding of abstract number magnitude, an importance which continues into adulthood (Moeller et al., 2011). Scores on a finger gnosis[13] test can predict maths ability (Fayol et al., 1998) and this can even be trained (Gracia-Bafalluy & Noel, 2008 although see Fischer, 2010). This points to a link between the use of fingers and long-term benefits to mathematical cognition which supports the thesis of embodied numerosity. Alongside this there is a growing literature that gesture is important to the understanding of mathematical concepts (Goldin-Meadow et al., 2009; Goldin-Meadow & Wagner, 2005) and there is evidence that children will give the right answer in gestural form before they can verbalise that answer suggesting a pivotal role for embodied knowledge (Goldin-Meadow et al., 1993).

[13] "The digital gnosia test mobilizes the ability to identify and organize afferents. Under visual supervision, the experimenter asks the participants to put their right hand in pronation and assigns a number from 1 to 5 to each of the participant's fingers, starting with the thumb. Participants then close their eyes, and the experimenter touches their fingers one by one in the sequence3–1–5–4–2. Participants must identify the touched finger by stating the associated number. The digital discrimination test requires a sensory and visuo-motor integration." (Fayol et al., 1998, p.B65)

Figure 1.2

A Histogram of Journal Articles and Published Proceedings Mentioning Embodied or Embodied Cognition from 1990 to 2019. Number of Publications Based on a Web of Science Search Conducted on the 12/07/2020 with the Key Words "Embodied" or "Embodied Cognition". The Number of Publications Relates Only to Journal Articles and Published Proceedings in the Research Area of Psychology.



Enactive Approaches: The Mind Moulded

Much like embodied cognition, enactivism does not maintain clear conceptual boundaries. I follow Loughlin (2020) in observing at this stage the position of enactivist cognition and the arguments for a third wave of extended cognition appear very similar. The extended mind third wave builds on Sutton’s (2006) complementarity thesis. This posits that the systems of interest are dynamic entities and proposes a non-individualistic agency that is not centred on the biological agent but instead distributed across people and things. This third wave of the extended mind argues against the idea of an agent centred cognitive extension (outlined above) and suggests instead that soft formed cognitive assemblies do not need to have organism centred cognition (Kirchhoff, 2012).

This view of an extended mind closely echoes arguments in distributed cognition where the locus of cognition is not centred on a human agent (Hollan et al., 2000; Hutchins, 1995). Additionally, to add to the confusion the foundational arguments for enactive cognition come from *The Embodied Mind* (Varela et al., 2016) and for a while it has been subsumed in the embodied cognition perspective. However, there is a distinct emphasis on the reciprocal nature of agency and distributed causality which is an important distinction even if this makes for a leaky theoretical vessel.

To lend a focus to the overview here, I shall draw on what enactive cognition can tell us about our primary research question: How people solve problems. Thompson (2016, pp. xxvi–xxvii) describes cognition in terms of sensemaking as:

> […] the exercise of skilful know-how in situated and embodied action. Cognitive structures and processes emerge from and constitutively depend on recurrent sensorimotor patterns of perception and action. […] a cognitive being's world is not a pre-specified, external realm, represented internally by its brain, but is rather a relational domain enacted or bought forth by that being in and through its mode of coupling with the environment.

Hutchins (2010a) emphasises the active nature of an enactive approach to cognition. It is not active in the manner of an omniscient homunculus recruiting external artefacts and yet maintaining her internal coherence and fixed properties, rather it is a dynamic and reciprocal process in which the properties of the external and internal are shaped by *action*. In that way, it makes no sense to disentangle the internal and external—on a macro and a micro scale our cognition is shaped by our actions and the environment which surrounds us. Cognition does not consist of Lego blocks, but rather Plasticine, which changes form through action.

Enactivism encompasses a wider cognitive ecosystem than all the other forms of extended cognition and is also more open to reciprocal relations. Gallagher (2017, p. 11) writes:

> The explanatory unit of cognition (perception, action etc.,) is not just the brain, or even two (or more) brains in the case of social cognition, but dynamic relations between organism and environment, or between two or more organisms, which include brains, but also include their own structural features that enable specific perception-action loops, which is turn effect statistical regularities that shape the structure and function of the nervous system.

It is this emphasis on a wide, entangled, reciprocal, polytemporal and heteroscalar cognition that marks the enactivist position. An enactive approach to numerical cognition, for example, sees mathematical activity as inherently connected to the material both in development and in the action of using the material to think. Artefacts and embodied forms structure human understanding of number as people create artefacts to strengthen their understanding so that the two co-evolve (Malafouris, 2013; Overmann, 2018).

Ecological Approaches: The Mind in Action

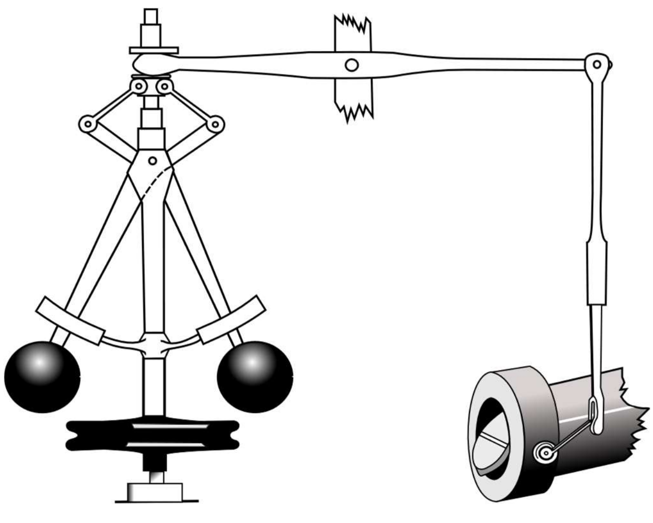
Ecological cognitive psychology[14] is closely related to the dynamic systems approach and assesses cognition in movement and action. A cornerstone paper in this regard was Van Gelder (1995) who used the example of Watt's flywheel governor (Figure 1.3) to suggest a dynamic systems approach to cognition in reply to Newell's suggestion that there "do not seem to be any viable alternatives" to a computational theory of mind (cited in Van Gelder, 1995, p.346). The cotton industry required uniform power to be delivered from the new steam engine but the rate of work and the pressure in the boiler

[14] Ecological psychology can cover many bases from work which addresses Neisser's (1967) call for ecological validity to Bronfenbrenner's (1979) theory of ecological systems. In this section I look solely at ecological psychology which is concerned with cognitive coordination with the environment.

were in constant flux so the engineering challenge was to create an instrument which was sensitive enough to adjust the valve to keep the pressure constant. Watt designed a centrifugal governor which spun according to the power being delivered by the steam engine. When the engine gave off too much power the arms of the flywheel spun fast enough to rise and close off the valve which reduced the power, the arms fell and the valve opened again (see Figure 1.3). Van Gelder used this to demonstrate that intelligent behaviour does not require a central computer[15] or representations. In this he was echoing ideas from Brooks (1991, p. 144): "We do claim however, that there need be no explicit representation of either the world or the intentions of the system to generate intelligent behaviors."

Figure 1.3

A Centrifugal Governor. The Steam from the Pump Increased the Speed of the Flywheel Which Caused the Balls to Fly Out and the Pump to Close Reducing the Speed. This Meant Steam Pumps Could Be Prevented From Overheating with No Human Attention or Central Computer. [16]



15 The use of computer here is to be understood in the broadest possible sense, that is something which computes.

16 Taken from https://commons.wikimedia.org/wiki/File:Centrifugal_governor.svg

Brooks had greatest success in designing ‘intelligent’ systems by rejecting the idea of a central intelligence and instead building on the work of William Grey Walter (Barrett, 2019) to demonstrate that intelligent behaviour can emerge when in contact with the environment without a need for an additional layer of abstraction. For example, Herbert[17] is a can collecting robot who navigates the world with a limited repertoire of simple behaviours that require no central representation or programming. Dynamic systems are, by definition, emergent; that is, the macro level of behaviour displays properties which cannot be explained by the properties of the constituent part and can only be explained as a whole (Hurley, 2001).

Ecological psychology builds on this work to explore how the affordances in an environment can render the need for a central processor obsolete. It is rooted in the work of Gibson (most famously Gibson, 1979) and considers the relationship between the world and the mind to be one of affordances (Chemero, 2003). An ecological psychologist posits that there is a direct link between perception and action. Cognitive systems should be thought of as “dynamic singularities” (Hurley, 2002, p. 2) with perception and action entwined in dynamic feedback loops: “The dynamic singularity is centred on the organism and moves through environments with the organism but has no shape boundaries”. For ecological psychologists the environmental stimuli are rich and form part of the cognitive ecosystem – they reject the classic cognitive sandwich which places cognition between perception (input) and action (output; Hurley, 2001).

The notion of affordances[18] is complex. The two basic positions are that affordances are either static properties of the environment or they exist in a dynamic

[17] Herbert in action can be seen here: https://youtu.be/YtNKuwiVYm0
[18] For all those externalists who decry the retreat to “representations” at the first sign of cognitive inexplicability, there are internalists who view “affordances” in the same way (e.g. Goldinger et al., 2016)

relation with the perceiving agent. I follow Chemero in viewing affordances as relations between the agent and her environment (Chemero, 2003, 2011). Simply put, an affordance describes a coupling between the agent in the world so that perception and action can be run without "thought". Affordance and affordance-agent interactions offer opportunities for action which then occur without the need for cognition. In this way, the ecological psychologists decentre the brain in cognition. They also, and importantly for the argument, bring the role of the material and extrabodily objects into the cognitive ecosystem.

For ecological psychologists, affordances can offer more parsimonious explanations of action. Take for example the outfielder ball problem. The problem that is posed here is how an outfielder in baseball catches a fly ball, that is a ball which is hit in such a manner that it flies up high before hitting the ground. This type of ball should be caught by the outfielder and remove the batter from play. The problem in this situation is that the ball is small and moving very fast and the player will need to carefully position themselves to be able to catch the ball. A computational model of the process would require the player to calculate the angle of the ball and move in an efficient and direct line; the complex computational requirements however soon overwhelm the computational power of the agent and belie the agent's quick and fluid movement. In turn, the data appear to support an ecological approach to the problem, that is the outfielder moves in coordination with the ball – an action[19] that might appear pragmatically inefficient but builds on the quick, tight and iterative coupling between perception and action. This leads to the agenct adopting a curved path towards the ball rather than the more direct line a computational model would predict. A complex action is explained far more

---

[19] This sort of action is best seen as an epistemic action (Kirsh & Maglio, 1994) where the pragmatic value is nil or even negative but the epistemic value is greater. This will be discussed more at length in Chapter Two.

parsimoniously by positing an action-perception link (Fink et al., 2009; Wilson & Golonka, 2013). As with performance on the matrix reasoning test, understanding of human intelligence and performance has been constrained by the models imposed which direct our attention inward.

Interactivity: The Mind in Coordination

Interactivity describes the coupling of the agent with her environment. It is commonly examined by allowing an experimental participant access to movable problem representations and assesses whether this leads to increased performance. It sits uneasily in any of the three categories so far outlined. On the one hand, the focus on cognitive offloading that can be seen in much of the work on the extended mind underlies the arguments for an augmentative effect and it holds to a computational approach where the environment is used when required. For example, the Tetris example in Clark and Chalmers (1998) was based on research conducted by Kirsh and Maglio (1994) which is a cornerstone text in the experimental work on interactivity. On the other hand, the theoretical writing on interactivity emphasises the reciprocal and dynamic nature of the coupling. Indeed, interactivity can be seen to straddle the enactive and the extended mind traditions with reference to both "reciprocal causations" and "changes in the environment in a way that suits agents' thinking needs […] to facilitate information processing" (G. Vallée-Tourangeau & Vallée-Tourangeau, 2017, p. 140). It has both a computational basis, items from the external world are often "recruited" and an ecological one—it is a form of "sense saturated coordination that contributes to human action" (Steffensen, 2017, p. 86).

It is this emphasis both on the environment as actively shaping and the agentic control of the environment that sets interactivity research in both extended and enactive traditions. Research in this interactivity straddles both cognitivist and non-cognitivist models of thinking and computational and non-computational models. G. Vallée-Tourangeau and Vallée-Tourangeau (2017; Figure 1.4) suggest updating the information processing model to include these external resources which classical models abandon on the other side of the human-environment interface. Their model is considered systemic in order to "underscore [their] view that cognition emerges from a complex set of entities (human and non-human) that form an interconnected network of reciprocal causations" (p. 142) in other words, a system. In this respect, this model commits itself to a kind of extended functionalism[20] in which the system is considered the cognitive unit. In this model (SysTM; Figure 1.4), external resources are not something to be ignored or tacked onto the end of cognitive architecture but are instead key constitutive parts of cognition. This model rejects linear processing and replaces it with processing loops which emphasise a reciprocal, non-linear nature. In the deductive loop, mental processing suggests actions which alter the environment and provide information which feeds back into the problem solver's understanding.

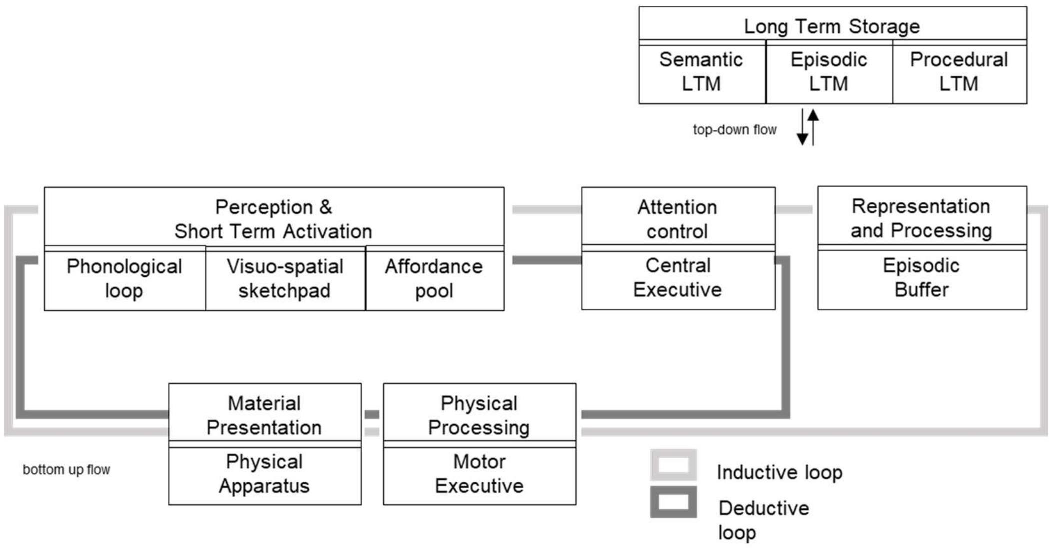
Crucially, the inductive loop in this model supports action that does not emanate from a mental plan but one that is guided by the environment. This places it more in the realm of enactive cognition. The addition of an affordances pool alongside the normal slave components of working memory (the phonological loop and the visual spatial sketchpad) guides the actions of the thinker without recourse to mental planning and allows the action to be generated from perception. Thus, this model posits that a problem

[20] The merits of extended functionalism are beyond the scope of this thesis but in brief, functionalism suggests that there is not anything which determines whether something is a mental state a priori but rather assesses its function within the cognitive system. The extended functionalist position (Wheeler, 2010) suggests that there is no one mark of the cognitive (Adams & Aizawa, 2001) which attaches to a thing but rather it is rather a role or function that carries a cognitive mark.

solution can emerge both deductively and inductively across the spatio-temporal trajectory of problem-solving.

Figure 1.4

The Systemic Thinking Model (SysTM) Adapted From Vallée-Tourangeau and Vallée-Tourangeau, 2017, p. 143.



The suggestion here is that problem-solving can be distributed across the cognitive resources of the problem solver and the material and social affordances yet also have a computational aspect. This allows for action to be part both of effortful cognition and also to flow directly from environment affordances. In this way it allows for the computational models and agent centric modelling of the extended mind hypotheses alongside an active and dynamic relationship with the environment.

Take, for example, Wilson and Clark's (2009) suggestion that the locus of cognition should not be the individual but rather, following Bingham's (1988) idea of task-specific devices, the object of investigation should be transient extended cognitive systems. These are soft-assembled cognitive ecosystems that mesh the problem-solving contributions of the human brain and central nervous system with those of the (rest of the) body and various elements of local cognitive architecture (Wilson & Clark, 2009). These can range in the way they recruit the environment from a simple dissipation of effort (such as the handle on a paint pot which facilitates carrying) to complex combinations of forces seen in something such as windsurfing. The soft assembled and transient systems do not exclude the biological brain and they alter with experience; for example, a new problem may require a new system which is discarded soon after use, whereas an oft encountered problem—such as a crossword—will lead to a well-rehearsed re-forming of a familiar set of resources. It is the emphasis on cognition emerging from a cognitive ecosystem (Steffensen, 2017) that has led to interactivity also being described as systemic cognition. There is a computational and algorithmic focus, but the computations are theorised to take place at the level of the system rather than the individual (Hutchins, 1995).

The initial focus of interactivity research was to contrast a low and high interactivity environment to track increased performance (see Maglio et al., 1999; F. Vallée-Tourangeau et al., 2016, among others). Research on mathematical cognition, for example, would track agents' performance when given movable numbered tokens as well as bringing into focus the relationship between interactivity and individual differences (Ross et al., 2020). This wave is much more in line with research in the tradition of the extended mind. However, a second wave of interactivity research has moved beyond viewing the environment as a cognitive extension to tracking the nature of the coupling and the resultant change in strategy that comes from being able to move and interact with the environment (Fleming & Maglio, 2015; Ross & Vallée-Tourangeau, 2020; Steffensen et al., 2016; F. Vallée-Tourangeau et al., 2020). This second wave with its emphasis on the material environment and the dynamic changes wrought to the agents by that material is much closer to an enactivist position. There has been a mirroring change in the underlying methodology to focus on the more granular process.

The Appropriate Level of Analysis

It is apparent from the multitude of positions outlined above, that the precise location of cognition is unclear and subject to movement according to the theoretical position being espoused. While plurality of approaches is to be welcomed, there are some levels of analysis that cannot comfortably co- exist from a metaphysical perspective. However, they make little difference from an empirical perspective. This tension means it is unlikely that the current stalemate will be easily resolved through empirical means because it is not an empirical argument (Kiverstein, 2018) and there is a danger of diverting empirical attention towards answering a question about what constitutes cognition which takes on a regressive circularity.

There is no doubt that the external environment influences cognitive processes. There is also no doubt that the old idea of disembodied brain has been called into question through the work outlined above and indeed, throughout its history. Even committed internalists such as Ohlsson propose externality volatility as a trigger for cognitive processes (Ohlsson, 2018). The real question is to what extent higher cognitive processes supervene on internal mental processes and to what extent they supervene on a wider cognitive system. This is not to deny the importance of the brain and internal cognition for cognitive processes rather to suggest that "alongside the physical brain, the organism's body and the world form part of the physical substrate that underlies behavior and cognition." (van Dijk et al., 2008, p. 298). In other words, it is trivially true to suggest that the environment causes internal cognitive change, what an externalist theory suggests is that not all cognition is mirrored in that internal change so that carving the problem-solving system at the skin level will exclude integral parts. If cognition is genuinely emergent then reductionism cannot work, if it is not and can be summed from an analysis of the individual parts then analysing these individual parts can still yield useful information.

Therefore, the selected level of analysis of a study reflects the researcher's underlying ontological commitments perhaps more than the methods used to probe that unit and any theoretical statements. Plato's reflections on the expert butcher "carving nature at its joints"[21] suggest that finding the natural weak spots in a concept is a more elegant way to proceed than hacking at a carcass. This has inspired philosophers to consider where conceptual joints lie, that is where concepts are easily distinguished one from the other. This is normally taken to be at the segments between natural kinds, that is naturally occurring, logical categories. The arguments about the existence or not of natural kinds and the possibility of identifying them is beyond the scope of this thesis (see Campbell et al., 2011 for an overview) but the psychologist is faced with finding the joints which separate causally relevant segments every time she decides on a focus of study.

This is not a trivial decision although it is often treated as such and decided by methodological norms and allegiances rather than the individual researcher. It is a naive experimental psychologist who believes that she can control or describe all exogenous variables, the question is where it is most appropriate to do so, where the links are weakest and least necessary (Hutchins, 2010b). If we believe that cognition consists of internal computations over mental representations, that it supervenes on internal states then it is a waste of time and resources to look elsewhere for the processes: An internalist and

[21] In Phaedrus (http://classics.mit.edu/Plato/phaedrus.html)

individualist research programme will yield all the answers we require. If we rather theorise that cognition is genuinely emergent from an interaction with things 'beyond the brain', then the reductionism inherent in most experimental psychology is a pale shadow of cognitive processes as enacted by embedded organisms and moreover will obscure important explanatory mechanisms. At best, an internalist research programme will stumble when scaling up problems, at worst it is measuring something which is simply an artifact of a stylised experimental situation and yet has perversely diverted a significant amount of scientists' time and resources (not to mention vast sums of public funding).

At its heart, the level of analysis a researcher selects reflects the hypothesised locus of the cognitive activity of interest. As Hutchins (2010b, p. 425) writes:

> At the level of organisms, bodies have different properties than organs, which have different properties than cells. In the realm of cognition, a neural circuit has different properties than any of the neurons in the circuit. The same can be said of a brain area with respect to the neural circuits that compose it, or of an entire brain with respect to the areas that interact within the brain. This is also true of the body/brain system with respect to either brain or body, and the world/body/brain system with respect to any of its parts. A system composed of a person in interaction with a cognitive artifact has different cognitive properties than those of the person alone.

Of course, it is possible for different units of analysis to coexist. That is, that narrow views of human behaviour and wide ones can both yield results which increase our understanding of a particular psychological phenomenon. Indeed, such multiplicity of perspectives is to be welcomed because it seems unlikely that any one viewpoint will be able to explain the complexity of processes that lead to intelligent behaviour. Furthermore, it is likely that the appropriate level of analysis will change according to the purpose of the enquiry. A pragmatic approach to the question of the level of analysis is described by Giere (2006) in relation to the composition of fluids. For the molecular chemist the appropriate level of analysis is different to the engineer who needs to model the flow of water through a system of pipes. That is why this thesis maintains a focus on the cognitive activity of interest. A plurality of approaches is not in itself antithetical to scientific progress, but each level must be meaningful.

I aim to demonstrate that an overly reductionist approach to problem-solving is not based on what people do when solving problems but is rather an artifact of a theoretical adherence to an overly intellectualised and internalist understanding of thinking. For thought in action (and I will argue that a large part of thought-full thinking is thought in action), this is not a *meaningful* level of analysis and risks missing out fundamental mechanisms which only emerge in interaction. Having reviewed the philosophical positions, I suggest that it is enough as practising researchers (rather than armchair explorers) that we make our decisions on the level of analysis based on choice between two explanatory models: Either one which holds that mental processes are self-sufficient and yield useful results when studied in isolation, or one which suggests that external props and events are important enough that not taking these into consideration is apt to miss the point. In this I follow Peirce (1878, p. 288) with the pragmatist's maxim:

> Consider what effects, that might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object.

Hochstein (2016) suggests that is counterproductive to insist that the domain of cognitive science can progress from a single conceptual framework whether that is internalist or externalist. I suggest that given the difficulty of empirically disentangling many of the positions outlined above, the framework that is used should not be determined by presupposed theoretical allegiances but what best suits the objects under investigation.

In that way, the underlying philosophical positioning is itself relational and contingent. I hope to demonstrate that problem-solving is a naturally embedded phenomenon and so increased understanding will flow from a situated perspective.

The Mind in Practice: A Pragmatist Approach to Cognition

This emphasis on "what works" for the situation draws heavily on the pragmatist approach. Pragmatism as a form of enquiry underlies much of what is discussed and argued in this thesis across several levels, from the position to be taken on the locus of cognition through the method of examining problem-solving and finally as an explanation for the behaviours witnessed. As Gallagher (2017) demonstrates many of the pragmatists' views on cognition foreshadow later views on embodied and enactive cognition such as Pierce (1868, p. 149) in a footnote suggesting that: "just as we say that a body in motion, and not that motion is in a body, we ought to say that we are in thought and not that thoughts are in us". The call for an action-based view of cognitive science has even been called "the pragmatic turn" (Engel et al., 2013). As this thesis takes a broadly enactivist position on the status of cognition, it is not surprising that there will be sympathies with the pragmatists running throughout.

However, the focus on pragmatism at this stage is not as an explanatory mechanism for cognition, I move away from a pragmatic stance on the ontology of cognition to consider it as an epistemological position on how to think about thinking. Broadly, pragmatism takes as its focus knowledge arising from action, from doing (Cornish & Gillespie, 2009) both on a micro level and on a broader scale. Thus, the rejection of metaphysical explanation for cognition applies to the position to be taken in this thesis: that the true way to find out about how people think and solve problems is through the act of observing nature, this not as a clue to a metaphysical truth but to describe the behaviours seen. Rorty (1991, p. 68) emphasises the recursive method of investigation under pragmatism: "new means changes ends, that you only know what you want after you've seen the results of your attempts to get what you once thought you wanted". Pragmatism focuses on the use of knowledge for action; under this position knowledge is generated for practical purpose. It is for this reason that it is the preferred philosophical position of mixed methods researchers who sit comfortably with apparent epistemological and axiological inconsistencies in pursuit of the best method to answer a research question (Onwuegbuzie & Leech, 2005, see also Chapter Three).

Problem-solving in the laboratory is the resolution of ignorance, scientific research is no different in nature. In this way, it is possible to adopt a pragmatist position on the way that thought unfolds in the world and to how we should best investigate that thought. For this reason, this thesis is less interested in the locus of cognition than what people do when they are cognizing and how the cognitive systems that facilitate that doing are formed.

Conclusion

The locus of cognition is unstable. This is to be expected when locus and ontology are enmeshed. It for this reason that I argue that an empirical programme should consider not the location of thought but how thought is manifest in the world. There are questions raised in this preliminary chapter which cannot be resolved empirically, however, a reductionist research programme committed to a modular and computational theory of mind risks missing out on the processes that occur in coordination with people and things because of an excessive focus on people and things. The empirical work presented in this thesis starts from the premise that problem-solving is something that is enacted in the world. It draws broadly from the traditions of research in enactivism and interactivity but applies a pragmatic lens. Other cognitive processes such as imagination, planning, daydreaming or mind wandering[22] are far harder to understand as distributed across the cranial boundary (see Goldinger et al., 2016 for a list of processes that resist easy externalist explanations) but problem-solving as we encounter it every day is *necessarily* situated: We solve problems generated by navigating an uncertain world.

I will demonstrate in the next chapter that the science of problem-solving is hampered by an excessive focus on mental and internal computations driven by an unexamined allegiance to an internalist perspective. Such a focus dictates the methodology employed to understand the behaviour and the process of problem-solving and in turn, the methods used make underlying assumptions about the ontology of the concept. If we assume problem-solving to be a matter for the intellect, then we will construct theories that do not allow for different explanatory mechanisms. This focus is problematic if we wish to uncover what happens when people solve problems rather than what problem-solving can tell us about ontologically unstable cognition. As Gallagher writes, "We, as minded beings are definitively 'out there', dynamically coupled to artefacts, tools, technologies, social practices and institutions that extend our cognitive processes." (Gallagher, 2017, p. 60).

The work presented here aims to demonstrate that drawing from different theoretical ideas of where to look for intelligent behaviour can expand the research into problem-solving as a process of moving through and resolving uncertainty. Allowing

[22] Although imagined things can often be nebulous until they take a form. Take for example Bardt's description of asking his art students to imagine an apple and draw it from imagination: "The results are remarkable for their consistency and cartoon-like flatness, each drawing a perfectly 'unimaginative' apple, almost indistinguishable from all the others. It turns out that the vaunted powers of the imagine disappoint when extracted from the mind as some kind of wilful mental operation. Only by either continuing the drawing or looking at and drawing a real apple does the imagination seem to stir or attention to phenomena re-emerge." (Bardt, 2019, p. 83)

cognition to reveal itself in action will make it more visible: “The self does not lurk somewhere between perceptual input and behavioural output, but reappears out in the open, embodied and embedded in the world” (Hurley, 2002, p. 3).

Chapter Two : Problem-solving

In this chapter, we move from a consideration of the ontological status of cognition as understood in the philosophical and psychological literature outlined in Chapter One to a closer look at the tool that will be used to measure cognition through the course of this thesis: Problem-solving. The chapter will first discuss the conceptual boundaries of problem-solving, its taxonomy and its importance to cognitive research before outlining models of problem-solving generated by classical, internalist views of psychology. I shall then examine the challenges from 4E cognition which consider problem-solving as necessarily engaged with the world. Finally, I shall discuss the limitations of using problem-solving tasks as a tool to measure problem-solving processes and offer some solutions to the problem.

Problem-solving

Problem-solving is manifest whenever a living creature navigates an uncertain external world. The problems faced can be mundane and demonstrably rooted in the resolution of uncomfortable physical states– how to get enough food to eat to remove the physical feeling of hunger – or far more abstract (yet still about the resolution of an uncomfortable state – think of philosophers wrestling with deep existential problems such as the nature of free will). Simply put, problem-solving in all its forms is an essential part of lived experience.

Problem-solving is also a foundational research area in the study of thinking and cognition. The study of thinking refers to effortful contemplation, the research area which also encompasses reasoning, and judgement and decision making. The boundaries between these disciplines are a useful but artificial construct; in practice there is considerable overlap between these aspects of thinking (Holyoak & Morrison, 2013). The three branches are separated as much by the tasks selected and underlying theoretical allegiances as a natural divide between them (Goel, 2010). There is a further hierarchical nature to our understanding of the type of problems that we choose to investigate under the banner of problem-solving: Abstract thinking is prioritised. We do not imagine Rodin's thinker to be contemplating how to turn a pile of leftovers into a healthy meal for example. This hierarchy is sustained by an enlightenment culture that privileges a recourse to unsullied human reason as separated from the needs and desires of the body (Lakoff & Johnson, 2010). The cognitive processes that underly this form of highly intellectualised *thinking* are by and large considered to be atomised and linear and marked by intentionality and rationality (De Jaegher, 2019; Pernu, 2017).

Types of Problem

Gilhooly (2019) suggests that all problems demonstrate three basic features: an end goal, a starting state and a set of possible actions which could transform the starting state into the end goal. Each of these areas is conceptually leaky and certainly the set of possible actions is not one which is easily constrained although the models do just that. Problem-solving researchers then classify these basic structures in three main ways: ill-defined or well-defined; knowledge rich or knowledge lean and insight or analytical problems.

Ill-defined and Well-defined Problems

A problem may have a well-defined or well-structured starting state, set of legal actions and end goal; in contrast, an ill-defined or ill-structured problem would leave any one of these three features underspecified. For example, the classic river crossing task presents problem solvers with a farmer on one side of a river with a fox, a chicken and a bag of grain. There is a boat which can take the farmer and one other piece of cargo. The problem requires the problem solver to move all three items across the river with the constraints that if the fox is left alone with the chicken on either side the fox will eat the chicken and similarly if the chicken is left alone with the grain that will be eaten. The task is well structured with each element clearly demarcated.

On the other hand, a problem such as how to combine a series of leftovers in the fridge to deal with various dietary requests is an ill-defined problem (although the operators are known to the problem solver). While the starting situation is clear, there is not a single normative answer against which to judge performance (Reed, 2016; Simon, 1977). Goel (2010) further argues that the constraints which define a well-structured problem are intrinsic to the problem itself and differ in type from other rules. For example, the problem of what to serve fussy family may be solved by turning away from the leftovers in the fridge and ordering a takeaway (or indeed, opting not to feed them at all!) Such answers would not threaten the integrity of the problem in a real situation[23]. However, opting to shoot the fox as a pest is not a permissible operator in the description of the river crossing problem. As we shall see in Chapter Six, the problems used in the laboratory are reliant on a mutual understanding of the "rules" and when this understanding collapses then the status of the problem also changes.

---

23 Although they would if the problem were presented in a psychologists' laboratory or indeed, a school classroom. Imagine a list of food in one column and a list of dietary requirements on the other and the task would be to satisfy all members of the family. The rules of the environment make it quite clear that the problem is to be assumed to be artificial. In this case the problem would be a well-structured one with a presumed optimal outcome. However, this is determined by the nature of the environment not the problem itself.

Knowledge Rich and Knowledge Lean

We can make another distinction between knowledge rich and knowledge lean problems. Knowledge rich problems are those such as chess or calculus that require a level of expertise to solve beyond the average well educated person[24]. A knowledge rich problem presupposes an existing mental schema that the problem solver can evoke as a framework when approaching the problem representation (Scheiter & Gerjets, 2003). For example, the problem of how to translate a text from a foreign language requires knowledge of both the content of the language and the skill of translation. In contrast, a knowledge lean problem assumes that no additional knowledge is required to solve the problem. Take for example, a problem such as the Tower of Hanoi. In this type of problem, the participant is required to move discs from one pile to another with various constraints depending on the problem environment. One does not need a level of additional knowledge or expertise to solve this problem[25]. It is important to note that this is a categorisation based on the characteristics of the problem but also the assumed characteristics of the problem solver. It makes specific and often implicit assumptions about what constitutes knowledge and what knowledge might be transferable or useful. Indeed, these assumptions about the epistemic and motivational state of the participants undergird much of the experimental work in this area.

---

[24] There is not a clear definition of what constitutes an average well educated person and while many of the research subjects end up being psychology undergraduates, there is a wide variation in these. Take for example the participant pool in Chapter Five who displayed a low level of anagram expertise compared to the participant in one of the papers used to design the stimuli. This is a neglected area in part because of the modular and nomological approach which seeks to flatten difference in search of universal laws.

[25] Although the participant does need the ability to understand the layers and layers of cultural and conversational pragmatics that turn her into a compliant interlocuter happy to play by the experimental rules. This knowledge is tacitly assumed within knowledge lean problems.

Analytical and Insight Problems

Two other main problem types have been identified which overlap with these other categories: analytical problems and insight problems. The problem categorisation here refers to the hypothesised process of problem solution. Analytical problems are those problems where the pathway to the solution is clear if at times arduous. The Tower of Hanoi described above is one such problem. These problems lend themselves to an easy path analysis and it is generally assumed that a normative answer was generated by a normative process (normativity here is defined in terms of optimality, namely the minimum number of moves required to traverse the problem space efficiently from start to finish).

Insight problems on the other hand require an abductive leap (Ross & Arfini, Forthcoming) – sometimes the goal state is already present (such as the triangle of coins problem) but the route to that end state is obscured so the route becomes the target for solution, other times the end state is not obvious and nor is the pathway to the solution. Take for example the 17 animals problem. This requires the problem solver to distribute 17 animals between 4 pens such that there is an odd number in each pen. The participant cannot know the end state or the way to get there; simple arithmetic rules are a distraction since no set of four odd numbers can sum to an odd number total (see Chapter Six for a longer discussion).

However, insight problems are also often simple, so the pathway and the end state are uncovered by the problem solver at the same time. It is not an insight problem if the participant can work out how to get to the answer but does not upon clearing the pathway stumble directly upon it. This means that analytical problems can progress by small moments of insight but here the insight is not attached to a correct answer but rather to a step in the progression towards that correct answer. This type of insight is not well explored by the literature in part because it is not something which is hypothesised to occur. An insight problem is supposed to require a sudden insight about the *solution*[26]. Sometimes this is accompanied by a clear phenomenological marker[27] (so called 'aha') and other times not. In other words, an insight problem is one which does not adhere to a typical normative model of problem-solving and requires different models. Insight problems struggle to be contained by models which rely on a series of rational steps precisely because they require the type of thinking which does not follow those steps (Ohlsson, 2018).

Despite an uncertainty about the process, we can draw clear distinctions about the nature of the ignorance faced by the problem solver. With analytical problems while the answer is unknown, the process of getting to the answer is clear and known. For example, with something such as mental arithmetic, while the problem solver may not know the solution, they will know the operators and rules that correspond to the most efficient process. These steps will have been already predetermined culturally and through personal experience. Therefore, it is likely that the agent will select the correct objects and the correct actions over those objects. Indeed, the accuracy of the solution in the eyes of the

26 Despite this, Webb et al. (2018) found the highest levels of insight for first anagrams and then the analytical problems in their norming studies.

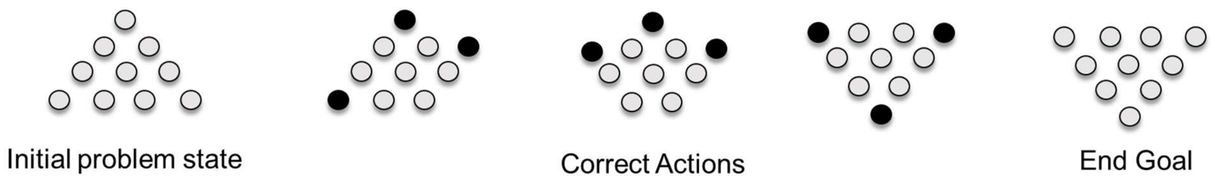
27 Initially insight problems were structured in such as way that it was presumed that they could only be solved through insightful processes. However a closer attention to the phenomenological markers by researchers such as Bowden (e.g. Bowden et al., 2005) and Danek (e.g. Danek et al., 2014) have led to these markers becoming the measure of 'insight' experienced and theoretically required to solve the problem shifting the process to the individual rather than as a necessary property of the problem itself (Webb et al., 2016; see also Chapter Three and Five). The phenomenology of insight is typically theorised as multifactorial and participants are requested to rate how they felt across five dimensions: happiness, certainty in the solution answer, surprise at coming across the answer, the suddenness of the answer and the level of impasse they experienced. A higher score represents a more insightful experience and is taken as an indication that the problem was solved using "insight" (Danek et al., 2020). It is not clear how the granular measure (participants are often asked to map feelings on scales of 0-100) maps onto a binary insight/not insight outcome.

problem solver is not clear in itself but is predicated on an accurate following of the steps. Trust in an answer to a mental arithmetic problem comes about not because of the answer itself but because we trust in the steps that led to the answer.[28]

For some insight problems, the goal state is given but it is the method of reaching the answer which is unclear. Take for example, the triangle of coins problem (e.g. F. Vallée-Tourangeau et al., 2020; Figure 2.1) which requires participants to move from one triangle shape to another in a restricted number of moves. This problem presents the problem solver with a triangle composed of circles[29] arranged in such a way that the triangle points up. The task is to change the orientation of the triangle so that it points down but by only moving three of the circles (Metcalfe, 1986). The answer requires the participant to move two coins up and the other one down as in Figure 2.1. The end state is clear and can be verified by checking against the target state.

Figure 2.1

The Triangle of Coins Problem and One Possible Solution Which Involves Moving the Bottom Vertices Up and the Top Vertex Down. [30].



[28] This is not to undermine the abilities of those who sense check whether an answer is correct or not but rather to emphasise the difference between a problem where the process becomes unimportant once the answers is generated because the answer is clear in itself and one where the answer is predicated on a correct, linear process.
[29] Rarely, despite the name, are participants presented with coins whether actual or pictures. The coins are abstracted into circles.
30 There are actually two ways of solving this problem. The first (translation across the median) is illustrated in Figure 2.1 and the second involves a rotation of the vertices while leaving the central rosette untouched, the rotational solution. For a fuller discussion of these two options see Vallée-Tourangeau et al. (2020).

To some extent this is the same with anagrams where the target word can be recognised by the problem solver once it is revealed. Thus it is not surprising that Chapters Four and Five report data which suggests that participants can accidentally generate the solution (see also F. Vallée-Tourangeau et al., 2020[31]). In this instance the process becomes unimportant, the answer is recognised as being correct and functions as its own accuracy check.

However, in many ill-structured or non-analytical problems, the problem solver is in double ignorance: They do not know the correct solution or the manner of approaching the correct solution. In terms of extended mind theories, this means that they will not necessarily select the correct objects or actions over them but merely those that satisfy the needs at the time (see Chapter Six). This echoes the Criterion of Satisfactory Progress Theory (MacGregor et al., 2001; Ormerod et al., 2013) which suggests that progress on insight problems requires satisfying local demands towards a predetermined criterion and posits a limited ability to look ahead. Incremental progress can lead away from the direct goal while fulfilling the demands of the current criterion[32]. There are clear examples of this are all four empirical chapters of the thesis.

While problem-solving research in the psychologist's laboratory looks at both analytical or insight problems, there is less of a focus on the ill-structured or knowledge-rich problems; the problems presented tend to be well structured and knowledge lean. Knowledge lean problems enable the testing of wide range of participants. Well-structured problems have normative solutions against which participants' "success" can be gauged. There are right answers and a field which approaches problem-solving from a rational and

---

[31] see https://osf.io/7nbkt/?view_only=4a3fe2db192f4dbabca859f7ce0e70e2 for video evidence.

[32] Note that these require a problem to have moves that can take you closer to or further away from the answer. As we shall see in Chapter Six, not all problems are structured in this way.

computational background provides normative methods of attaining these answers even if dual process theories (Kahneman, 2012) suggest that these answers can be the result of conscious or unconscious thought processes. This assumes that the problem-solving processes can be modelled and tracked using models of rationality. [33]

This thesis will look at four different examples of problem-solving. The first three studies report experiments in problem-solving which take place in a psychologist's laboratory: a word production task (Chapter Four), an anagram solving task (Chapter Five) and an insight problem (the socks problem[34]; Chapter Six). Chapter Seven moves to a different problem environment: namely a creative artist working with a difficult material. The first two studies follow traditional in laboratory design and employ the method of kinenoetic analysis which will be fully developed in Chapter Three and Chapter Eight. In contrast, the third bridges traditional research and emplys a quasi-ethnographic approach to laboratory research while the final study is a detailed case study using focused ethnography. However, all explore how agents solve problems.

All four studies start from the proposition that whether or not the relationship between the extra and intra cranial processes are causal or constitutive as discussed in Chapter One, a close examination of the nature of that relationship is necessary to understand problem-solving as it unfolds in the laboratory or outside of it. This externalist focus is as yet not part of the mainstream approach to problem-solving which adopts a far more sequestered approach because it assumes "thought full" cognition is internal in

---

33 It also assumes that a correct answer is arrived at using a correct process. Even when a problem is well defined, it is possible to arrive at the correct answer via a circuitous or even wrong route. For example, as often happens in maths, a 'double' error occurs where the second error undoes the first and the correct solution results. (I am indebted to Susan Cooper for this observation)

34 This problem requires participants to solve the problem: "You have socks in a bag in a ratio of 4 socks for every 5 socks, what is the minimum number of socks you would need to pull to guarantee a pair"

nature. I will start by outlining the classic models of problem-solving before moving to assess challenges to the internalist model.

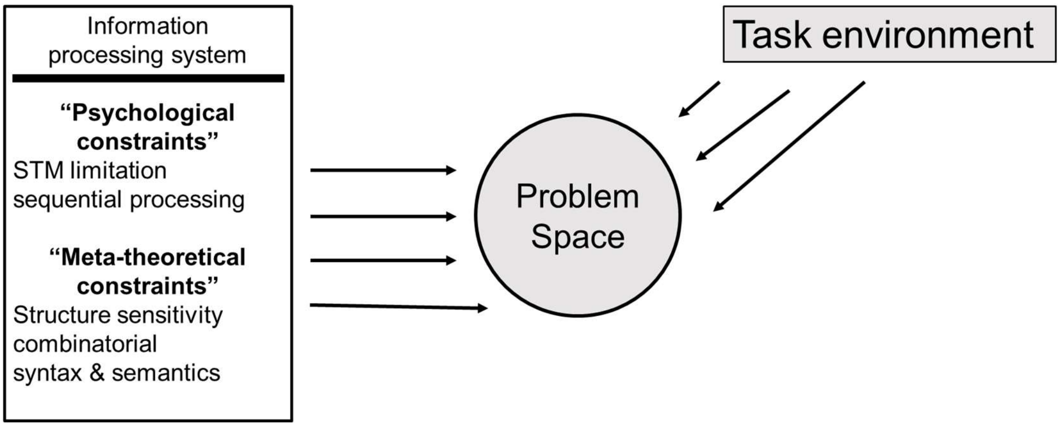
Classic Approaches to Problem-solving

The science of problem-solving has its roots firmly in an internalist conception of cognition. As outlined in Chapter One, the very nature of a higher cognitive process assumes that thinking as separate from action (See Footnote 1, p1). Higher cognition is what we do to initiate a planned action and is therefore separate from that action. This notion adheres to the classic cognitive sandwich (Hurley, 2001) separating thought and action. The environmental framing of the problem is deemed unimportant. More than that, it may even be a distraction. Everything which is necessary for the problem solution is contained within the person and the capacity to solve problems reflects psychometric properties of the person and the characteristics of the underlying algorithms presupposed by the problem. It is often these characteristics which are manipulated. This is all in perfect accordance with the view outlined in Chapter One that the human cognitive system is an information processing system which can be programmed much like a robot (Hollnagel, 2007).

The theoretical background to classic cognitive approaches to problem-solving draws on the information processing framework. This stipulates three main areas of investigation: the information processing system, the problem space and the task environment (Goel, 2010; Figure 2.2).

Figure 2.2

The Classic Approach to Problem-Solving Which Suggests a Problem's Characteristics and the Characteristics of the Problem Solver Come Together in the Problem Space and Dictate Problem Solving Capacity (Adapted from Goel, 2010; STM = Short Term Memory.)



The task environment under the classical framework refers not to a physical environment but rather the fundamental structure of the problem that is being solved. This task environment is not related to the physical instantiation of a problem. Take for example the classic triangle of coins problem (Figure 2.1). According to this theory and in line with the principles of methodological solipsism (Chapter One), the task environment remains the same whether it is presented as a word problem, a pictorial representation, a physical task with actual coins or a touchscreen version allowing changes in the array. The different physical contexts are irrelevant to an analysis of the task environment, the task environment is rather a "theoretical projection" onto a concrete situation (Kirsh, 2009, p. 266)

The information processing system is the agent doing the problem-solving. That Holyoak and Morrison (2013, p. 12) can suggest that "thinking [does] not necessarily require a human or even a sentient being" while also requiring a symbolic, computational representation demonstrates how tightly this proposal is bound up with research on artificial intelligence. Alongside the initial constraint of the fundamental limitations of processing power (working memory, for example), Goel indicates that the classical view also requires that such a system operates in a certain way, that is as a system which is able to perform computations over representation, computations reflecting a combinatorial syntax. In this he draws from Fodor and Pylyshyn (1988) who are unequivocal in the view that a classical view of cognition is defined by these two aspects: Cognition is computation over content bearing representations. It is this that allows the somewhat contradictory position that non sentient things can be cognize even while things are not considered cognitive. Cognition is the process of computations over representations whoever enacts it.

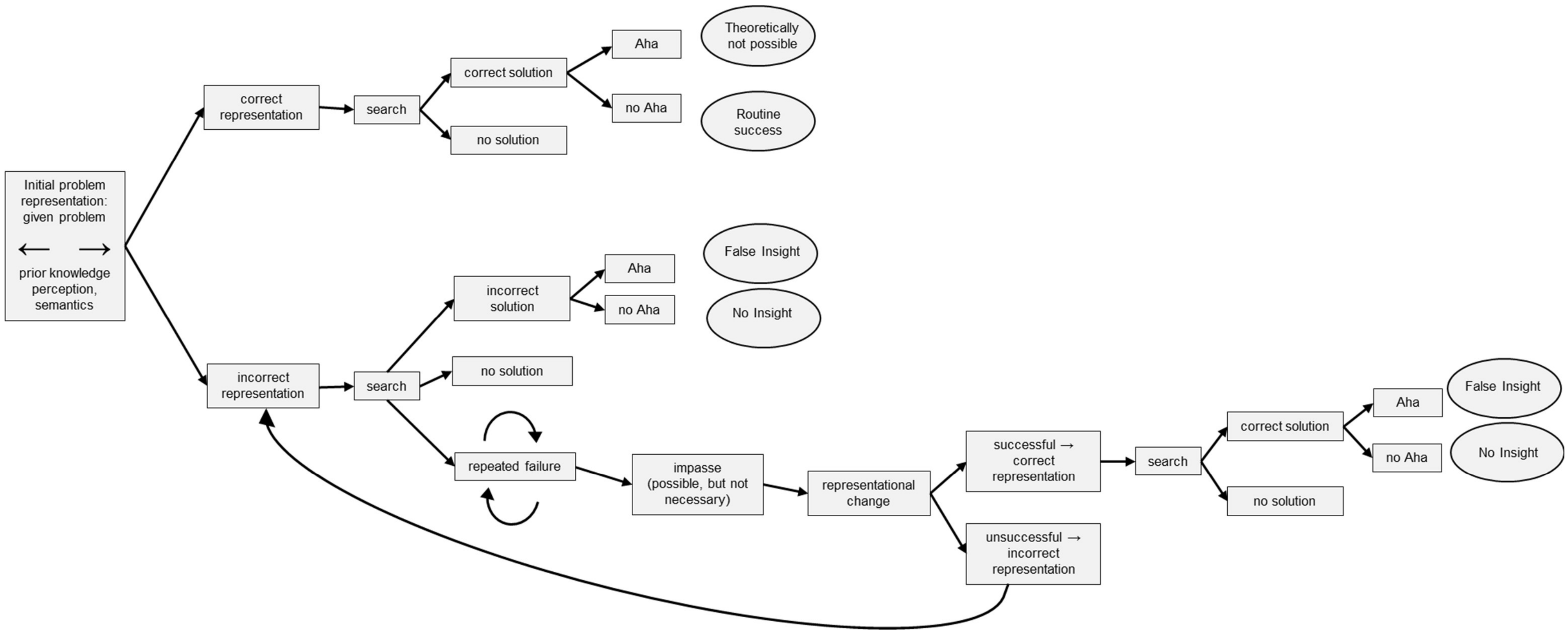
The problem space, therefore, is where these dual constraints of task environment and information processing system interact, and problem-solving research is committed to mapping this space; the interest is in how the manipulations of the task environment interact with the constraints of the processor. Take for example, work on mental arithmetic and working memory. Calculations (the task environment) are held constant while the capacity of working memory is either measured or manipulated by using dual task procedures which aim to suppress either the central executive or one of the slave systems of working memory (either the phonological loop or the visuo spatial sketchpad; see Raghubar et al., [2010] for a review of relevant studies). Performance in light of these constraints can be assessed as a function of the constraints because all else is held equal. Again, the use of space here is metaphorical, this is not an analysis of the actual physical space occupied by a problem solution. While it is possible for such a physical space to exist and for problem solutions to be traced through gestural or material representations, this is not the space of interest here. Rather this problem-solving space is the same as the problem environment: a hypothesised area in which the required representations are processed (Kirsh, 2009).

Let us look at this approach in more detail. Take for example the model proposed by Danek (2018; Figure 2.3). In this model Danek explores the possible routes to solution of a problem. Danek separates these routes into two different types: monotonic (upper pathway) in which the problem solver starts with a representation of the problem which does not need to be fundamentally changed and non-montonic (lower pathway), where a change in the initial problem representation is needed for a solution to occur. This model is inspired by the Representational Change Theory (RCT) outlined by Ohlsson (1984a, 1992). In short, RCT suggests that insight problems require that the problem solver gets stuck—experiences an impasse—and is required to restructure the initial problematic representation of the problem. This theory is discussed at more length in Chapter Five.

Note that this model adheres to the principles of a normative model of problem-solving with binary outcomes. The initial problem representation is either correct or incorrect and the success or otherwise of the problem-solving depends on this first step. There are no degrees here of the type of representation, even later in the non-monotonic arm which follows an incorrect representation. Yet, take, for example the triangle of coins problem introduced earlier (Figure 2.2). It is possible for a participant to understand that one vertex must be moved up without realising that the other two must be moved down. Similarly, the 'aha' feeling is either present or absent in this model despite the fact that it is now recommended that it is measured as a gradual process (Danek et al., 2014). The search is not considered in any depth – it is simply a stage to get through which happens presumably according to a normative process. There is some iterativity but it is largely flattened by the linear process. There is also no learning from an incorrect representation,

Figure 2.3

A Model to Differentiate Between Monotonic and Non-Monotic Problem-solving (adapted from Danek [2018])



you return to the same epistemic state after each incorrect representation. There is no Groundhog Day style redemption and or gradual change in the state of the problem solver built into the model.

What is possibly the most important aspect—the nature and quality of the search process—is truncated and assessed only in terms of its final outcome. Success or otherwise is predicated on the representation (theoretically internal in this instance) rather than the process. However, despite having a correct representation, there is no guarantee of generating a correct solution. There is little consideration in this model of how a correct representation could lead to no solution although this is a possibility in half the monotonic outcomes and half of the non-monotonic options even after the problem-solver has finally hit on the correct problem representation. This raises the interesting question of how a participant could know if their representation were correct or incorrect. Take for example, the case of an agent who has formed a correct representation but fails to offer a solution (note the model does not allow for an incorrect answer with a correct representation – an incorrect answer only exists as "a kind of exit strategy" [Danek, 2018] in the incorrect representation category) whether before or after insight. It is unclear from the model how that agent differs in terms of her epistemic state from the agent who has an incorrect representation but also fails to offer an answer.

Additionally, the model raises the question of false 'aha', a phenomenon which suggests that there must be a level of procedural error and there is no guarantee that generating a solution even accompanied by 'aha' is because of a correct representation. Procedural error can also lead to the correct solutions from an incorrect problem representation. It is not clear from the model how you would know whether a participant had a correct or incorrect representation from the outcome aside from whether the solution was correct or incorrect. A lack of solution could indicate an initial correct or incorrect representation.

If problem-solving is seen as a change in the epistemic state of the participant, as moving from a state of ignorance to a state of understanding then it becomes clear that we need to assess (a) the original epistemic state (b) track the process of state change and (c) the final state. Current problem-solving research rests on the assumption that the original epistemic state is the same for all people (beyond a binary one determined by whether they have encountered the problem before or not), that the process of state change is the same and can be determined by the answer and also that the final state is binary: You have either solved the problem or not and a correct solution indicates a correct problem representation. Given the circularity here—a correct problem representation is the only way of getting a correct answer and the problem representation can only be measured by a tool calibrated by eliciting a correct representation— arguably places research in problem-solving at the risk of lapsing in a unprofitable state of regress.

Challenges to the classic approach

The Locus of Representation: Reifying the Problem Space

Not all problem-solving research assumes that the problem environment and problem space are metaphorical and interchangeable across different material presentations. It is true that the tight conceptual boundaries which define problem-solving in terms of computations over internal representations do not invite a more externalist approach but this definition is not a necessary facet of problem-solving research. Even working within this paradigm of representations and operations, the material environment of problem-solving is gradually being taken into account.

This move of the problem space from an abstract one to one with an actual physical form reflects a move from a second order problem space to a first order one (F. Vallée-Tourangeau & March, 2019; see also discussions in Chapter Six which looks at this in more depth). The problem space in most problem-solving research is second order, that is to say it is presented in such a way as to only allow an abstracted interaction with it. In this way it requires a mental representation be formed because there is no other way to proceed. Take for example the triangle of coins: When the problem is given only as a picture on paper which offers no chance to rearrange the tokens, participants have no choice but to rely on their internal resources and skills in mental rotation. Again, this is the outcome of a methodologically solipsistic research programme which sees no difference in the problem environment because the internal representations are able to be treated in a discrete and modular way. On this view, rotating in the world translates to the same underlying cognitive process as rotating in the mind so there is no need to look beyond these internal representations. Of course, as argued in by F. Vallée-Tourangeau and March (2019), the reliance on mental representations necessarily leads to evidence for the importance of these representations: There is no other way of solving the problem.

Contrast this with first order problem-solving where the same problem is presented in such a way that it can be rearranged (F. Vallée-Tourangeau et al., 2020). It is likely that such a presentation will evoke different processes to mental only ones. Kirsh and Maglio (1994) argue that actions on the world can have an epistemic rather than pragmatic function. Using the example of a Tetris player spinning a zoid, the authors demonstrate how the classic information processing model cannot account for the way that actions generate knowledge. Instead of a model of action which requires planning before action, the data suggest that rotations in the world act to improve the epistemic state of the player by reducing “the pace, time or unreliability of the computations” (p. 527). These actions change the input to the processing system in much the same way that we have seen that a fielder catching a fly ball moves herself to change the input to the processing system (Chapter One). From the experimental data, Kirsh and Maglio extract two sorts of actions that can be performed on the array: Pragmatic actions which are movements towards a goal and epistemic actions which make no practical advances in the problem space but reveal information and so advance the epistemic state of the participant. The divide between the two is perhaps best illustrated by a navigator taking a boat around an island[35]. To navigate successfully around the island requires hugging close to the coastline, this is a pragmatic action. However, to better get a sense of positioning relative to landmarks on the island, the navigator may sail further out to sea. This is an epistemic action, it holds no practical purpose, indeed, quite the opposite as the boat moves away from its goal, but it advances by changing the epistemic landscape. Kirsh and Maglio point to the number of superfluous moves made by Tetris players which did not advance the players pragmatically but did advance them epistemically to sustain the idea that action can yield information.

Experimental research in interactivity has a strong focus on the manipulability of external representations (perhaps because of the overall commitment in the first wave of research to cognitivist computational models). It traditionally contrasts low interactivity environments with high interactive environments. Participants are presented problems which invite and allow an interaction with the environment typically in the form of a movable external representation such as numbered tokens (Ross et al., 2020 or letter tiles (Maglio et al., 1999). However, people can also be invited to interact with more complex

35 I borrow this example from Andy Clark in a talk given on June 17th 2020 (see https://blogs.ed.ac.uk/virtualppig/ for details)

physical objects such as string and thumb tacks (Chuderski, et al., 2020) or even water jars and water (F. Vallée-Tourangeau et al., 2011). These artefacts can be actual artefacts (e.g., Fioratou & Cowley, 2009) or representations which can be moved on a tablet (e.g., F. Vallée-Tourangeau et al., 2020). The only constraint is that they must offer some level of engagement; the array needs to be able to be altered to facilitate problem-solving. Necessarily these are first order problems.

In a high interactivity condition, the participants are invited to move the artefacts as they choose. In the low interactivity condition, they are restrained from moving the artefacts. The nature of the low interactivity condition varies slightly from being given pencil and paper (Chuderski, et al., 2020)[36], to be allowed to move freely but with static artefacts (Chapter Four and Five) to having movement restrained and hands laid flat on the table (Vallée-Tourangeau et al., 2016). The current research programme into interactivity has been established with the main aim of demonstrating the augmentative effect of cognitive extension. An experiment has 'worked' if it establishes that performance is better in the high interactivity condition. It has had some success with this although it seems likely that there is a publication bias at play (those studies which do not show a benefit of interactivity will be less likely to be pursued by the authors) and the evidence reported is inconclusive at times with interactivity certainly not a panacea in all circumstances and across all levels of individual differences or tasks (Chuderski et al., 2020; Maglio et al., 1999; F. Vallée-Tourangeau & Wrightman, 2010; Webb & F. Vallée-Tourangeau, 2009)

36 In this study the authors gave their participants a pen and pencil in the main study. This was an unusual choice when such things obviously function as external aids. In a second, smaller study they removed the scaffold from participants (see Chapter Six).

Insight vs Analytical Problem-solving: Equifinality of Outcome.

Problem-solving in practice does not adhere to one normative and linear model so different models are proposed and tested. Following the principle of rationality outlined above, models of problem-solving are based on normative ways of solving the problem. However, some problems are solved in different ways and so modifications must be made to the hypothesised architecture.

Traditional views on problem-solving have assumed the process based on the normative models constructed above and the binary nature of the problem-solving outcome supports or does not support hypotheses based on these normative models. The assumption has been that those who were successful followed the model and those that were unsuccessful did not follow the model or followed it poorly. On this view, the model plays the role of an explanation but without a check to assess whether that model tracks real time behaviour. Moreover, the dichotomy between the two types of problems is based on an a priori decision of how the problem should be solved (Bowden & Jung-Beeman, 2003).

At times this is problematic. Take for example, anagrams. It is possible to solve anagrams through trial and error, and indeed this would be the way a computational model of problem-solving would be likely to predict a solution occurring. After all, for each five-letter anagram[37] there are only a maximum of 125 possible options. The cognitive skills required to work through 125 possible combinations are not taxing and although motivation would be crucial and people may require a pen and piece of paper to scaffold working memory and systematic exploration, it would not be beyond the ability of most.

37 5 -letter anagrams are typical, although there is some variation in the number of letters used in the anagrams by researchers: Ellis et al. (2011) use four letter anagrams with a distractor letter; Kounios et al. (2009) used four and five letter anagrams; Valueva et al. (2016) use slightly more with stimuli ranging in length from five to seven letters; Webb et al. (2016) use five letters.

This has led to them being classed until recently as analytical problems (Gilhooly & Murphy, 2005). However, it seems likely that they can also be solved in an insightful way. The first indication of this is the time taken to solve an anagram – some anagrams are solved in under 2 seconds (Novick & Sherman, 2003) and it is unlikely that this happens through computational methods (although this complicated by the evidence that these 'pop out' solutions are more likely to occur when participants have some level of expertise.)

The other evidence for different ways of solving the problem comes from reports from participants themselves. Indeed, recent normative research indicates that the anagrams not only elicit reliably high ratings of insight (Webb et al., 2018) but do so more than classic insight problems. In fact, the research from Webb and colleagues across two studies coupled with other work from Fleck and Weisberg (2013) suggest that the binary split between analytical and insight problems may be misplaced. Rather than assuming insight from the problem, it appears that there are many different phenomenological responses elicited by the correct solution. It remains to be seen if these different responses relate to different underlying cognitive processes (although see Danek et al., 2020). The equifinality evident in participant reports challenges the suggestion that a correct solution can only be produced in a single way. Models which ignore participant behaviour except tracking a binary correct or incorrect outcome run the risk of collapsing when they encounter reality outside of the psychologist's textbook.

The recent challenge to the split between analytical and insight problems revealed the inadequacy of the underlying and unexpressed theoretical background. A logical model was generated to explain how the problem was solved. This model was predicated on a closed system and a normative path: those who got the correct answer were presumed to have followed this logical model. It is only through introspective, first person and semi-qualitative methods that this has been shown to be a poor reflection of how actual problem-solving occurs. Indeed, the large-scale examination of how problems were solved undertaken by Webb et al. (2018) was motivated because “it just seemed clear from participant reports but could not be seen in the data” (Webb, private correspondence). Until Bowden and Jung-Beeman (2003), a correct solution was assumed to have been generated in the same way and this was determined by the task environment. This is an unhelpful artifact of using an algorithmic and linear model of problem-solving. As seen above the correct solution has a narrow path leading to it under such a framework. Such a theory does not have space for other ways of solving problems. However, a move away from a priori models, opens the opportunity for a variety of process which may more accurately reflect the data from the more qualitative surveys.

Webb et al. (2018) assessed the self-reported levels of insight across a variety of classic insight, classic non insight problems, anagrams and compound remote associates[38]. They found that self-reported insight was highest for anagrams followed by compound remote associates, then classic non insight problems and then classic insight problems. This suggests that the affective component of insight may not have such a strong link to the insight problem type. It remains to be seen with clarity what the link is between problem type, the affective dimension and the cognitive dimension are.

It was only through eliciting phenomenological reports based on introspection that the models were challenged. Indeed, while the field has now added insight measures to most new tasks, it has done so without a true appreciation of the challenge that this poses to the models of problem-solving. Given that the same type of introspection has

38 Compound remote associates present the participant with three words which can all be combined with the same target word to produce a valid word such a: RIVER, SAND and ACCOUNT with can all be combined with the word BANK.

downgraded the status of impasse from an essential component of insight problem-solving to something that may or may not occur, it appears that models which are based on rational supposition rather than real time behaviour may not map onto problem-solving as it actually unfolds. Tracking where the models overlap with behaviour and where they do not is essential.

Problem-solving as Instrument: The Dangers of Experimenter's Regress

Much of the argument of this thesis relies on accepting the psychologist's laboratory as a very real environment and the experimental process as worthy of study in itself rather than just as a source of neutral data. Hacking (1982, p. 34) argues that psychology is not a laboratory science because "there is too little of that 'apparatus used in isolation to interfere"'. However, whether or not it counts as a mature laboratory science (and indeed whether it ever could under the radical terms outlined by Hacking because of the necessary substantial ethical constraints), if the underlying philosophy of experiments in science overall is under investigated (Radder, 2003) then in psychology it certainly suffers from an extreme form of theoretical inattention: For example, while there are entries on experiments in physics and biology in the Stanford Encyclopedia of Philosophy, there is no partner entry on experiments in psychology.

This neglect is not unproblematic. One of the problems posed is the risk of experimenter's regress which psychology with its focus on latent variables would seem to be particularly vulnerable. The idea of experimenter's regress was first introduced in Collins (1975) in which he argued that the enculturation of knowledge is important to scientific experimentation and that knowledge could not be transferred algorithmically so that the only way of knowing that someone had the skills to replicate an experiment would be for them to replicate the experiment successfully. This leads to regress: a tool functions as its own test. The example he gives in a later publication is designing an apparatus to locate gravitational waves to establish their existence:

> What the correct outcome [to the question of the existence or not of gravitational waves] is depends upon whether there are gravity waves hitting the Earth in detectable fluxes. To find this out we must build 'a good gravity wave detector and have a look. 'But we won't know if we have built a good detector until we have tried it and obtained the correct outcome! But we don't know what the correct outcome is until ... and so on *ad infinitum.* (Collins, 1985, p. 84)

For Collins, the only way out of this is to find a way of fixing the apparatus outside of the experimental situation. This becomes more difficult when we are dealing with psychological traits that have no underlying physical substrate such as intelligence or personality and which can only be measured by the behaviours which they are also theorised to explain. This is something which Boag (2015, p. 3) describes as “verbal magic” where entities are invented to describe behaviours which are then taken as evidence of the invented entity. Boag’s target is the research field of personality but Sugarman (2017) makes a similar observation about psychology as a whole. He refers to this as “psychologism”. The basic process of psychology that he outlines is that a behaviour is observed, and this behaviour is assumed to have an internal motivation. An instrument is designed to measure that behaviour and in so doing reifies it. In all cases, there is a case of experimenter’s regress because there is no validation for the measuring instruments aside from the behaviour and there is no way of measuring the behaviour aside from the instruments. As Sugarman notes, this process is particularly difficult in psychology where the participants are active in their own study so the instruments may create the phenomena they are measuring, trapping the researcher in a bottomless and unproductive regress.

We can see this in the use of problem-solving as an instrument to test cognition. While it is hard to find a clear and consistent definition of higher order cognitive processes (see Footnote 1, p1), problem-solving constantly features as a process which falls into this category (De Jaegher, 2019). There is a self-referential circularity: The process and method of testing the process are collapsed. We test higher order cognition through observing how people solve problems because people use higher cognition to solve problems. The problems to be solved are artificial – knowledge lean and well defined. In addition, performance is therefore easily aggregated across multiple tasks and this performance is hypothesised to reflect an underlying cognitive architecture independent of the individual (Ohlsson, 2011). Indeed, this aggregation is encouraged especially in the case of "highly idiosyncratic" (Chuderski et al., 2020, p. 5) insight problems. Once this underlying cognitive architecture is identified it can, presumably be applied across the different problems, the environmental relief flattened in terms of unidimensional computations.

However, problem-solving is also the object of interest, so the tool is designed to measure itself which as we have seen above can be problematic. There are also underlying theoretical assumptions that constrain the nature of the tool: namely the nature of the thinking required. Duncker writes that a problem requires effortful thinking (Duncker, 1945) but this in an untested and circular assumption. Effortful thinking is measured through the use of problems because problems require effortful thinking. Ecological psychologists have had much success collapsing this binary split between action/thought which is the premise of much problem-solving research. In this field, there is an investigation of what might be thought of as "lower order" problem-solving such as the outfielder problem in baseball (Fink et al., 2009; see Chapter One). Indeed, the skilled international framework (Rietveld et al., 2018) expressly sets out to "dissolve the dichotomy between 'lower cognition' and 'higher cognition'" (p.43). However, such problem-solving is not the sort of problem-solving that the science of problem-solving is aiming to capture.

A science of problem-solving assumes that if a problem can be solved without conscious contemplation or "thought-full" ness then it is not a problem of interest because problem-solving is primarily a tool to uncover the structure of thought. The field of problem-solving arose precisely as a tool to eliciting certain types of contemplative behaviours (Simon, 1965). The conceptual boundaries of a 'problem' are therefore contingent and yet determine the ontology of what is under investigation. This ontological understanding is further constrained by and determines methodological allegiances. It is here that we run the very real risk of experimenters' regress: Establishing the cognitive processes required to solve problems uses a tool for measuring those processes which only deemed a success if it elicits the processes it is required to test.

Conclusion

Under the classical model of problem-solving differences in the material environment are assumed to be unimportant as are differences in the potential process; the first is irrelevant and the second unconsidered. This runs the very real risk of encountering experimenter's regress in which the evidence for the process is generated only by something which assumes the existence of that process to function. In the next chapter, I shall suggest a novel way of anchoring the process models to real behaviours

Chapter Three : The Relationship Between Ontology, Epistemology and Methodology

Chapter One outlined the different theoretical approaches to the underlying concept that this thesis aims to investigate, namely cognition. Chapter Two took a critical look the tool that has been selected to research cognition, problem-solving. This thesis takes the position that cognition is best viewed as emergent from an open system, so this chapter will explore methodologies for exploring this system. Current research in problem-solving and cognitive psychology more generally is bound to methods which rely on a methodological individualism (G. Vallée-Tourangeau & Vallée-Tourangeau, 2017), my aim in this chapter is to suggest other ways of conducting research on cognition. It is these methods alongside the underlying concepts that will be refined over the course of the thesis.

I will begin by taking a close look at some of the assumptions which underlie experimental research, a method which I follow in the first three studies reported in this thesis. I follow Andersen et al., (2019) in arguing that there are fundamental philosophical biases to any scientific research that need to be examined to progress knowledge and devote the time here to unpick the biases in experimental research. I shall then demonstrate how the methodology underlying much research in problem-solving reflects its own theoretical commitments about the ontology of both the problems and the cognitive processes. I suggest that these theoretical commitments can lead to experimenter's regress and that the restrictive models create blind spots where important explanatory factors are not taken into consideration.

I will go on to propose a mixed methods pragmatic approach which takes the form of systematic progress tracing using the objects which also function as experimental stimuli: kinenoetic analysis. Such an approach allows a researcher to track the dynamic trajectory of problem-solving, the formation of cognitive systems and the systemic, idiosyncratic nature of embedded and enacted problem-solving. It will also allow us to start to fill in the detail which may have been missed in existing models. As the research here is mixed methods and combines both qualitative and quantitative paradigms and epistemological outlooks as well as relying on observations that I have generated, I offer some brief reflections on the role of constructivism and subjectivity which are explored more fully in Chapter Eight.

Cognitive psychology has a difficult challenge. It sets out to measure something which is at best ontologically unstable (as demonstrated by the plurality of approaches outlined in Chapter One). The object of interest cannot be seen or directly measured it, a researcher can only measure physical traces (whether neural, material or behavioural) which stand in for cognition and from which she conjures the mechanisms which drive those traces. It is for this reason that the argument for the ontological locus of cognition has yet to be decisively articulated. It has no material reality which arguments or empirical experimentation will uncover. Therefore, it is recreated every time it is examined. Cognition is located not in the world nor in the head nor in some in between, rather cognition can only be demonstrably found in the tools designed to measure it. This chapter examines the nature of the methods currently used and what they reveal about the often-unstated ontological allegiances. I will then go on to describe the tools that will also be used through the course of the thesis. [39]

[39] This is necessarily an introduction to the method used. This thesis covers a lot of ground and several experimental methods. Each chapter relies on different methods and experimental stimuli so this chapter will only lay out the overall theoretical justifications for taking an observational and mixed methods approach. Further discussion of the measures and tools employed proceed the chapter in which they are used Chapter Eight will synthesise the findings and methodological advances of this thesis.

Classic Research in Cognitive Psychology

Cognitive psychology is dominated by a positivist research tradition which tends towards studies of a quantitative nature (Bickhard, 1992; Kuiken et al., 1992; Smythe, 1992). The positivist position is rarely explicitly stated but as Bickard (1992, p.1) argues it is "implicit in the organisation of thinking" and a necessary outcome of the stated aim of uncovering the nature of the computational representations which drive human thought.

There are many philosophical trends in the debate around positivism but for the purposes here, I shall define it as a an epistemological position that value free knowledge can be produced by empirical methods to yield universal and generalisable laws (Breen & Darlaston-Jones, 2010) [40]. The foundational aim of cognitive science was to uncover the nature of these laws. It therefore assumes an object about which a progressive research programme can uncover a hitherto hidden truth, in this case about human cognitive processes. By aligning itself with such a position, cognitive psychology was able to step away from the purely material observations of behaviourism to "open the black box" (Gigerenzer, 2019, p.4) of unseen processes while maintaining the ideals of the scientific method and the dominant methodology of experimental research (Goel, 2019).

An internalist and computational approach to cognition starts by trying to uncover and elucidate the internal factors which produce intelligent behaviour whether these are processes or psychometric properties. The focus is therefore, necessarily on the internal processes. As in Chapter One, the environment is seen as complicating efforts to map these processes. Cognitive psychology progresses therefore from a top-down position which decides on the capacity to be explained and posits a series of relations between input and

[40] Hacking (1982, p. 43) suggest that "no-one today wants to be called a positivist" but there are no value judgements attached to positivism in this case.

outputs that would explain that capacity before testing that hypothesis through a series of elegant "if…then…" suppositions. This research programme is underpinned by the assumption that thinking unfolds much like a programmable algorithm (Cummins, 2000). As seen in previous chapters, such a question constrains the possible answers and is possibly a poor reflection of intelligent behaviour more generally. From a practical perspective, it posits as the object of interest something which is fundamentally unobservable.

Such an ontological perspective dictates the methods used to examine the underlying processes: Elegant experiments are constructed to test different theoretical models. Experiments anchor the "fragile science" (Wilson, 2001) of psychology to the harder sciences and help to satisfy the so called "physics envy". Experiments are considered key to the scientific method because they "disentangle variables and test them in isolation, that they use comparison and justify results through replication, [and …] they exclude, through blind or double-blind designs, experimenter bias and subjective expectations." (Knorr Cetina, 1992, p.114). The extent to which this can be achieved in psychological experiments is curiously underexamined.

The dominance of experimental research and quantitative methods in cognitive psychology is sustained by epistemological circles (Elder-Vass, 2012). These circles can be considered as disciplinary boundary objects that stabilise relative ontologies and allow their investigation in a scientific manner. Often, therefore, the justification for a method is anchored by the epistemological circle. To those outside this circle at times these justifications can seem bizarre – take for example creativity research in which the number of uses participants can generate for a paper clip is taken as a direct proxy for creativity – but the important thing to note is that they are justified beliefs within the context of the epistemological circle. On this view scientific knowledge is produced in interaction and is constrained by methodological and epistemological norms. It therefore behoves any serious researcher to take a closer examination of those norms[41].

Cognitive psychology is dominated by the experimental method. Hacking (1982 p.75) describes experimentation as interference in the course of nature" and from this argues that the experimenter must have some belief in the ontological reality of what she is manipulating. The experimental manipulation requires a realist position at the moment of experimenting. It is a tool designed to reveal through elegant design an aspect of the world which has hitherto remained veiled. This role of an experiment, as something done to elicit a certain behaviour so that behaviour can be examined is perhaps far closer to the real nature of experimenting in psychology rather than the lip service paid to the falsifying demands of null hypothesis significance testing ("the null ritual"; Gigerenzer, 2018).

The restrictions on how the course of nature is to be diverted in psychology are

---

[41] The methodological norms of quantitative psychological research are not fixed and the lack of reflexivity over these changes is at times puzzling especially in the quantitative sciences (there is a much clearer embrace of the necessary reflexivity of the research process in qualitative research). As Gigerenzer (2019) points out there is a bidirectional relationship between method and technology or new statistical techniques The 'tools to theories' heuristic he describes demonstrates the enmeshed nature of theory and tools for discovery. I have already outlined the relationship between the different forms of technology and the metaphors extracted to describe thinking. Gigerenzer goes further describing how, for example, attribution theory relies on the analysis of variance statistical procedure that has been developed just previous or how factor analysis enabled research into factor models of personality. Beyond a heuristic for understanding and generating theories about human behaviour, technological and statistical development constrain and build methods which feedback to shape methodologies.
Take for example the recent emphasis on statistical power and complex models: This requires larger numbers of participants and the software necessary to conduct the analyses. This has been facilitated by platforms such as MTurk or Prolific where participants can be recruited for a small cost to perform experiments via a virtual platform and by the use of free, opensource software to facilitate these analyses. This supports advances in statistical power and techniques which are sustained by the relative ease of achieving this marker of reliability. These methodological choices are not neutral. For example, this reliance on online participants directs the ontology of a "problem" by restricting problem types to well defined and knowledge lean problems described in Chapter Two. This is further sustained by the need for clear and easily analysed performance measures such as the proportion of correct trials or the latency to solution. The distance involved means that there are no way of measuring other factors that haven't already been accounted for. This sustains a false security that all additional factors have been controlled

tighter than in experiments in other scientific disciplines. These restrictions stem from clear ethical concerns but also because of the reflexivity of the psychological situation. Not only can the objects of experiments in psychology not be subjected to extreme unethical procedures but they are often interested in the experiment itself (and if they are keen, psychology students will take a keener interest). The truism that the laboratory is messy (Peterson, 2016), is more important when you posit a system that is in a dynamic interaction with that messiness. There is a greater chance of this affecting the nature of the outcome. The question becomes twofold: Is it possible to control for that messiness and is that control even desirable?[42]

This is not to reject experiments. Experiments allow the psychologists to get their hands dirty and test their models against reality. They are the pragmatists' method of avoiding experimenter's regress by generating situations and observing cognition in action. However, the nature of an experiment in psychology does need to be carefully examined. I argue that when an open system is being assessed, an experiment can no longer function as the hypothetico-deductive tool but rather has a different ontology and purpose. It acts to

[42] It is interesting in this respect that the practice in a psychologist's laboratory has garnered little attention in the aim to increase the amount of reproducible psychological science. An extensive search on laboratory practices yielded only one (Brenninkmeijer et al., 2019). This paper from 2019 states that "unwritten [lab] practices have not been systematically studied; the current work represents a first attempt in this direction (p.1)" and indeed the reference section is heavily reliant on research published in the last century and blog posts. There is only one paper cited from after the replication crisis which deals with infant research. This is a curious situation especially as one of the people who has been most critical of the reproducibility movement, Baumeister, makes it quite clear that his effect requires a "motivated researcher" (Baumeister, 2019, p. 5) to elicit it. One can only speculate on why the method of designing experiments and analysing data has attracted so much attention but the method of collecting it has not. There is an implicit hierarchy in science between the technicians and the scientists, between theory and practice (Hacking, 1982) which means that the hands on nature of data collection is often not considered and is farmed out to undergraduate volunteer RAs, the fruit flies experimenting on the fruit flies. Whatever the reason, there is a lack of focus on behaviour in and out of the lab that must have consequences for our understanding of that behaviour. Cleaner and cleaner data is sought because there is no way of knowing what has driven the outliers. Of course, this can lead to the situation that the average participant in psychology is so average that meaningful difference is erased. This has led to experimenters in medical trial reintroducing those participants that would have been excluded as too complex in randomised controlled trials (Anjum et al., 2020) in the form of pragmatic controlled trial.

create a situation to first allow us to describe behaviours. This inductive method allows us to draw patterns from complexity. It also recognises that there is an inherent constructivism in our understanding of quantitative data (see Chapter Eight for a longer discussion).

Adopting the perspective outlined in Chapter One, requires us to re-examine the traditional methodology which is reliant on a modular view of cognition. Mental states simply cannot be understood without references to the environment in which they are situated. A methodology which is reliant on closed systems is inappropriate for measuring the dynamic and idiosyncratic open systems which are in a state of flux and depend on a complex interaction between the properties of both the person and the environment. If the argument that cognitive states are best measured at the level of the system rather than the individual is accepted, then tracking the moment the system forms and its performance when it is formed becomes essential to understanding the underlying systemic processes and necessitates a new set of methodological approaches, a methodological interactionism (F. Vallée-Tourangeau & Vallée-Tourangeau, 2020).

The purpose of this overview is to demonstrate that while experimental research can be useful to elicit the behaviour of interest, it relies on epistemological underpinnings which mean that the data generated cannot be separated from those and nor from the beliefs about the ontology of the subject of investigation. This is as true for the work reported in this thesis as for the work reviewed. It assumes the object of interest is an open cognitive system and therefore the current methods of experimental psychology which are predicated on a closed system require examination.

A Mixed Methods Approach

I suggest that the answer to the problem of how to study open and transient cognitive systems using experiments (which are normally a method that relies on a closed system) is to reconfigure the ontology of an experiment. Current approaches in experimental psychology cannot challenge an internalist viewpoint because they are predicated on that very unexamined perspective (F. Vallée-Tourangeau & March, 2019). The models of rational and linear problem-solving assumes process through performance, rather than through tracking behaviour and so may not provide satisfactory explanations of cognition which does not follow the computational model. A detailed programme of observation will allow us to track explanatory factors that have been neglected. Experimentation can be retained but as a way of stirring nature to provide the behaviours which are of interest. Rather than the outcomes being subsumed into existing theories and being used to measure adherence to that process, I suggest adopting a 'bottom up' approach. This will help us to focus our exploration on "how does it work?"(Cummins, 2000) rather than searching for nomological laws of behaviour. In this I follow Martin and Bateman '(1993, p. 3) who suggest: "perfect knowledge of how many times each letter of the alphabet recurs on this page would give no indication of the text's meaning." I shall further demonstrate that alongside quantitative data, qualitative data can be analysed at both the level of the group and the individual. These qualitative data take the form of video and visual data rather than text generated data and thus require observational techniques rather than interpretivist ones[43].

A Pragmatic Approach to Methodology

Most research is mono-methodic, selecting either quantitative or qualitative approaches and most sub disciplines in cognitive science have a quantitative dominant approach (Mandler, 2011). Each of these positions has its own underlying epistemological philosophy and makes assumptions about the ontology of what is being researched. Quantitative research is associated with a more positivist view point and uses the methodology of the natural sciences to test hypotheses whereas qualitative research has as a focus the understanding of the participants and is based in a phenomenological or interpretivist tradition (Bryman, 1984). Methods have been subsumed under the philosophical position and the contents of the methods become less important than the research philosophy which underpins them (Niglas, 2010). Thus self-report measure of, for example, the phenomenological reports of insight are collected and aggregated and quantitivised even though they are based on introspective measures. Bryman (2007) reports that mixed methods researchers struggled to integrate research from both traditions because they view the two areas of research as distinct paradigms with distinct epistemological allegiances.

For many researchers this difficulty reflects the incompatibility thesis (Bryman, 1984), that is the suggestion that the different approaches cannot possibly be merged

[43] The epistemological stance here is nuanced. Qualitative data do not always imply a qualitative epistemology and quantitative data are also constructed. In other words, the research reported in this thesis does not take its epistemological position based on the nature of the data but rather from the nature of the analysis and the subsequent interpretation. There is nothing less constructed about data which are distilled in the form of numbers than those which are distilled in the form of behavioural codes. Chapter Eight expands on this in more depth.

because they have different underlying epistemologies. The so called ‘paradigm wars’ led Rossman and Wilson (1985) to describe three major schools of thought: Purists, situationists and pragmatists . Purists suggest that the complete disconnect between the ontologic, epistemologic and axiologic assumptions of qualitative and quantitative research mean that it is illogical to ever mix them. Situationalists are somewhat less extreme although they hold to a binary disciplinary divide which assigns some questions as more suitable for investigation by each set of methods. Finally, pragmatists doubt the underlying divide extends to the level of methods, rather they argue it stops at the interpretation of the data collected. The methods are epistemologically neutral and are often mixed without a conscious consideration of the underlying research paradigm. The work presented in this thesis adopts this pragmatic approach[44]

Qualitative and Process Based Approaches to Problem-solving

A research programme which profiles binary outcomes and perhaps some psychometrics can tell us how successful certain people can be in problem-solving, but it can only speculate on the reasons for why. This can be seen in the insight problem-solving research programme; while still dominated by a quantitative research, some of the more important findings recently have come from more qualitative approaches. These finding have cast doubt on the idea of a pure insight sequence which can be extracted from aggregated scores whether that is on a procedural or phenomenological level. It is important to note that these qualitative approaches offered the only way that the pure insight sequence rhetoric could have been challenged. Prior to this, certain aspects were

[44] The paradigm wars have somewhat died down since the 1980s and more nuanced epistemological positions are now far more the norm (Creswell & Plano Clark, 2018)

assumed because they were built into the model and hence were sheltered from critical reflections. For example, an insight problem was designed in such a way to create an impasse, therefore those that solved it were deemed to have encountered an impasse. However, there is converging evidence that points to equifinality in the problem-solving solution and this needs to be considered in building predictive models. The qualitative research to date has taken three main forms: verbal protocols and self-reports and granular process analysis through eye tracking or other measures. Here I will briefly review these before moving on to illustrate how qualitative research will be used in this thesis.

Verbal protocols

Verbal protocols are perhaps the most common way of tracking process. They consist of asking participants to speak while they are performing a task, in this case solving a problem. There are two main problems with this, one theoretical and one practical. First, cognitive psychology posits the presence of unconscious processes which are beyond understanding (Ball & Ormerod, 2017) and so reliance on an individual's in-the-moment introspection may mean that important processes and explanatory factors that occur outside of the problem solver's conscious awareness go unnoticed.[45] Therefore, it is a technique for understanding the information which participants are paying attention to rather than all the contributory factors. Second, the use of these think aloud techniques may have overshadowing effects, that is to say that the task of speaking while problem-solving may change the process although this is unclear (Fleck & Weisberg, 2013). Despite this, think aloud protocols have many benefits for understanding procedure and for fleshing out

[45] In a series of papers on serendipity (Ross, 2020; Ross forthcoming a, b), I have argued that it is exactly this framing of serendipity through the experience of the person rather than through observation which has shaped the current view of serendipity as an experience rather than an event and means that we miss moments which are not filtered through our experience.

process beyond analysis of latencies or binary performance outcomes. For example, Fleck and Weisberg's (2013) study employed verbal protocols specifically to test the times when a "pure" insight sequence occurs. The data they report from this more granular, process-based method suggests that the nature of insight is more complex and diverse than existing models (e.g., Ohlsson, 1992; 2011) may assume.

Self-report

The other evidence for different ways of solving the problem comes from reports collected after the task and relate to the feelings elicited on realizing the solution to the problem. When it comes to insight, insight is considered to have a distinct phenomenological marker, corresponding to the impasse resolution stage. Research from Webb and colleagues across two studies using these post task reports (Webb et al., 2016, 2018) now suggest that the binary split between analytical and insight problems, a key theoretical foundation stone, may be misplaced. Rather, much as illustrated by verbal protocol data outline above, insight is idiosyncratic and unreliable and is not guaranteed by a certain class of problem.

Theoretical models of the cognitive processes underlying insight problems have been updated after behaviour and self- report is taken into account; the moment of impasse has been downgraded and there is an increasing appreciation that problems are solved in diverse ways which cannot be easily predicted from the problem itself. When it comes to phenomenological approaches to insight, it is still unclear that this adds to an explanation of process. The distinction between phenomenology and process is being collapsed and problems which generate higher self-report scores of insight are said to be solved more insightfully but as evidence emerges that propensity to experience insight may be an important individual difference (Webb et al., 2019), the link between phenomenology and process needs to be established rather than assumed (although see also Danek et al., 2020).

Process Analysis

When it is employed, finely grained analysis of behaviour displays clear evidence of equifinality, that is multiple pathways to the same solution. This same point is clearly made by Steffensen (2016) as he reflects on the variability demonstrated in video analysis of an insight problem reported in F. Vallée-Tourangeau et al. (2016). The same masking of individual pathways by aggregate scores is also reported in Cushen and Wiley (2012) who used participant reports of the important elements in solving the triangle of coins problem to track the moments when they realised that the vertices were the critical components. The (Chapter Two, Figure 2.1) triangle of coins problem can only be solved by leaving the central rosette untouched and focusing on the corner coins. Tracking the moment of realisation of the importance of these coins indicate whether the realisation came about gradually or incrementally, and a comparison could be made between a cued and a non-cued condition. A subsequent individual analysis indicated that while behaviour in the main supported the conclusions from the aggregated means that the cued condition showed incremental insight, there were cases of individuals displaying sudden restructuring which were obscured by aggregate means.

Bilalić et al., (2019) found similar evidence for diverse pathways to solution. They used eye tracking to monitor participants' problem-solving of matchstick arithmetic tasks. The task presented participants with incorrect equations in the form of matchsticks and asks them to move a certain number of matchsticks to make the equation correct. The original equation is VI = VI+ VI. Here the answer requires the problem solver to shift their focus from the numbers to the operators so that the second addition sign can be turned to another equals sign yielding the (tautologous) answer VI=VI=VI. The critical moment of restructuring comes when the participant shifts their focus from the roman numerals to the operators. Crucially, Bilalić et al. report that the pattern of eye movements for some indicated a gradual understanding of the problem solution and for others a more sudden one and that indeed that these cognitive dimensions of insight interact with the affective ones: Unsurprisingly, the participants who showed an incremental approach reported lower ratings of surprise and suddenness.

Material Traces as a Research Tool

A research programme which is predicated on an understanding of cognition as behaviour must pay attention to the actual behaviour which is manifested. Furthermore, the binary split between mind and behaviour an 'environmentalist' approach is predicated on an internalist and Cartesian perspective. If rather the position that the mind manifests itself in behaviour, and that a dualistic split between the two is unsustainable is accepted then to measure behaviour is to measure the mind. Thoughts that occur in coordination with objects can be reconfigured as object-thoughts. Measuring behaviour is also particularly important for research in interactivity. This requires the assessment of behaviour both as a manipulation check (experimental conditions are meaningless if they do not result in an actual difference in behaviour as described in Maglio et al., [1999][46]) but also as a measure of the main topic of interest.

I argue therefore that the outcome to a problem solution is not simply an output from a series of inputs and mental operations but actually manifests as behaviour. A

[46] In this paper (which is described in more detail in Chapter Four) the authors let us know that around 1/3 of the participants in the high interactivity condition chose not to interact with the artefacts.

problem is generated by a situation and solved in that situation. There is no *a priori* reason why the outcome variable should be solution rate other than an adherence to the computational model of the mind which uses speed as a clue to the underlying processes. Increasing the range of dependent variables available to understand how problem-solving unfolds can only refine and add to the research programme.[47] The process tracing in the studies above were mapped by either self-report or by eye tracking. Through this thesis, I shall be demonstrating that it is possible to track the process of thinking through the material traces left by participants in a high interactivity problem-solving environment. Such a technique was used by Christensen and Friis-Olivarius (2020) who asked participants to brainstorm on sticky post it notes and used these external traces to make inferences about the internal traces. As Christensen and Friis-Olivarius (2020) argue, the use of movable objects allow us to track thoughts through action without the need for expensive technology.

It also reflects the theoretical position elucidated in Chapter One that these movements in the world do not reflect thought but also constitute it, so that through object-movement thought is unveiled. Thus it is similar to the research programme in gesture (Goldin-Meadow et al., 1993; Hostetter, 2011; Hostetter & Alibali, 2019) which looks at gesture as both reflecting and generating thought. Under this object based analytical framework, the researcher's knowledge in these cases comes from the movement of object-thoughts in the world rather than from a contemplation either of cognitive process abstracted from the world or self-constructed life worlds. Both these approaches return the focus internally, to what is happening inside the problem solver (Latour, 1999). The data here are not

[47] This is not to be confused with researcher's degrees of freedom and *p* hacking (Simmons et al., 2011). As I demonstrate in Chapter Five, an inductive approach to data can be done in a principled manner.

generated by text-based analysis but by observations of behaviours and actions. These observations come from a mapping of the physical nature of these transformations of and actions on objects in space and time.

This is not a perfect method of tracking thought. As Bocanegra et al. (2019) have demonstrated there is a rhythm to thought in the world. Thoughts are enacted and generated in a dynamic relationship between the internal and the external. Tracking the material traces of thought make some parts visible but leave other occluded. This is problematic for some researchers. Ohlsson suggests that "the fundamental mistake of the environmentalist approach is that it replaces the goal of explaining mind with the goal of explaining behavior" (2011, p.27) and continues to mischaracterise the externalist position as being determined to place causality wholly in the environment. Simon makes a clear separation between "cognition" and "behaviour in general" (1969, p.53). This similar self-evident split is taken up much later by Aizawa (2017) who uses the example of patients with locked in syndrome who still display different brain functions in response to different stimuli to demonstrate that cognition is not the same as behaviour. I disagree with Ohlsson that the approach here is misguided by focusing on behaviour. The work presented in this thesis is not interested in explaining behaviour either through recourse to hypothesised cognitive structures or to constructed life worlds, the primary research aim is to track behaviour to unpick how systems form to solve problems. When mind is generated by behaviours then the distinction between the two are collapsed. Those behaviours are visible and generate something measurable yet are currently underexplored.

Given the clear circularity that is endemic to problem-solving —a correct problem representation is the only way of getting a correct answer and the problem representation can only be measured by a tool calibrated by eliciting a correct representation—a return to observation, what people do, is required to ground the research and avoid regress. Relying on the aggregate measures will obscure a simple truth about creativity: it is not in the person, nor in the world, but rather in their co-constitutive coupling.

Kinenoetic Analysis

The method of kinenoetic analysis (KA; Ross & Vallée-Tourangeau, under review) I employ in the first two studies (Chapter Four and Five) reported in this thesis relies on the material traces of thought enacted in the world analysed pluralistically. In the third study (Chapter Six), I explore the limits of this method. In short KA suggests that through actions on objects, thought becomes visible and thought processes become manifest. Furthermore, KA unveils previously hidden explanatory factors by removing the need for a participant to be aware of those factors. It focuses on the observable and verifiable actions on objects. This detailed attention to environment chance and complexity and moving the focus away from the psychometric properties of the person, has important implications for the design of environments to maximise problem-solving success[48]. This type of deeper and more finely grained analysis becomes particularly important when there is less control over aspects of the experimental set up. This is especially salient each time participants are embedded in a complex, materially rich world. Many explanatory aspects of problem-solving go unnoticed when behaviours are only observed through a theoretical lens. KA focuses on tracking changes in the world, changes in objects, rather than the conjectured changes in the creator's mental representation of these objects. The form of KA described

[48] Although in this second wave of interactivity research, the scaffolding effect which is important as outlined in Chapter One, becomes of secondary importance.

here is made possible by the detailed scrutiny of video data and the instrumentalization of the problem task so that it becomes both measure and instrument.

Extrospection

Alongside quantitative analyses of behaviour, this thesis uses finely grained observational data to support and explain the quantitative outcomes. This allows an analysis of behaviour across large data sets. Qualitative analyses are turned into quantitative data via objectively generated codes. The work presented in these parts of the analysis draws from the methodological basis of interaction analysis. Interaction analysis refers to the systematic analysis of behaviour through observation (Bakeman & Gottman, 2009; Bakeman & Quera, 2011). It normally requires video data, but it can happen in situ with a researcher taking notes. The coding schemes can be generated in two ways: through an iterative process which allows themes to be generated through repeated watching of the data or by pre-specified coding to support prior hypotheses. Often the two ways can inform each other with a first round of coding using preformed codes also yield observations which can be categorised as exploratory. This use of codes to generate quantitative data from observations is central to the approach – it is systematic and quantitative rather than qualitative (Bakeman & Gottman, 2009). Objectivity is assured either by the use of clearly defined coding schemes – such as the onset of a certain clearly defined behaviour – or the use of two coders for behaviour which is more qualitative in nature. Extrospection can support an understanding of process outside of the subjective feeling of the participant.

A level of subjectivity remains in the researcher not only in the coding but the choice of behaviours to code. These behaviours can be decided before coding and the video coding applied using easily reproduced codes (as in Chapter Four and Five) or they may be generated inductively. However, at no times over the course of this thesis were behaviour codes preregistered and as primary researcher, I observed the behaviour of each participant before returning to the video data. Therefore, each code reflected observations generated during the experiment as well as theoretically motivated observations. This was unavoidable. I collected all the data for the research reported here and my theoretical allegiances are necessarily reflected in that data. However, the nature of extrospective data is that the codes reflect objectively verifiable moves on the world. Rather than asking participants to reflect on their inner experience, extrospective research lays open movements in the world.

Cognitive Case Studies

Current research in psychology is dominated by a Galtonian nomothetic model where the laws are based on the average of a sample (Robinson, 2011). The work presented here challenges some of the established scientific practice which requires both replicability and the possibility to predict events (Cummins, 2000; Neves-Pereira, 2019) by drawing some inferences from single cases. Predictive causality is hard to infer from single cases and while there are strong patterns of behaviour which inform the work presented in this thesis, it would require much more work to make the argument for a predictive science based on these observations. However, this does not mean the descriptions outlined throughout the thesis are either non-scientific nor non- informative. I follow Sutton (2010) in suggesting that a narrative and non-predictive frameworks do not diminish rigour. As Robinson (2011, p. 36) argues, the idiographic approach to scientific research is an approach to understanding a single thing:

> Science is not just the development and testing of theory, it is also the endeavour to describe and explain objects and events. Events, by definition, only happen once, and objects, by definition, are singular.

The work presented in this thesis is idiographic because it aims to generate data about particular events but in so doing also make broader comments about the process of problem-solving. There is a danger in reducing the epistemic credibility of science to two criteria which reflect only one research question and only one point in the research cycle. The results here maintain validity not through statistical analysis but by rigorous and documented methodological expertise (Leonelli, 2018).

In adopting this approach, I follow Steffensen (2016; see also Steffensen et al., 2016) who has suggested that cognitive psychology could benefit from the application of the probatonic principle. That is, a principle which focuses on the “single sheep that has our full attention and which is not reducible to being part of the herd” (p. 30). The argument for such a shift in focus is that the unique cognitive system which coalesces around each problem-solving agent expresses a form of variability which traditional analysis which focuses only on a binary correct or incorrect answer will inevitably not be able to identify. As seen in Chapter Two this equifinality can obscure important processes by collapsing different problem-solving methods into one.

Steffensen argues that there is no principled way of knowing a priori what the important moments of variability might be rather “to identify stable patterns, one has to investigate (the trajectory of) the cognitive probatonics of individual agent-environment systems” (p.31). These systems echo Wilson and Clark's (2009) transient extended cognitive systems which were described in Chapter One. Each time a participant enters an experimental situation, a system forms which is dynamic, fluid and idiosyncratic. This is true even if the researcher treats the system as a closed one. This is a willing blindness on the side of the researcher. As I have already discussed, rather than isolated, closed systems, every experimental situation is an open one with more or less permeable boundaries (Chapter Six is a particularly clear examination of the permeability of the boundaries). Quantitative research necessarily prespecifies the questions it will ask before observing the data. This allows it to test prespecified hypotheses but variables which are not already identified will necessarily be excluded.

Unlike the coded interaction analysis outlined above, the case study stage of the research reported in this thesis is intended to demonstrate existence not incidence (Smith et al., 1995) although it could inspire a new set of coding schemes or even future experimental research[49]. This phase of data analysis is purely qualitative and reduces the sample size to generate key critical themes that can be evidenced by behaviours and are therefore grounded in the data. There is a necessary subjectivity on the part of the researcher at this stage, however, the role of such an analysis is not interpretative but rather descriptive. Much like the extrospective measures employed above the interpretative stance is limited

One way a detailed granular analysis of particular participants can add to our understanding of a phenomenon is to assess whether the explanatory mechanisms assumed in the experimental manipulations are actually those which hypothesised beforehand rather than assuming them from the outcome. Thus, this deeper level of analysis straddles the observational and the experimental. It also does not derive its validity from positing causal explanations but rather by suggesting the mechanisms through which the causal

[49] Although this is not the primary aim. Description is sufficient.

explanations already established by the experimental results are realised. It may be that the hypothesised explanatory factors map easily identifiable behaviours in the case study material or it may be that other factors emerge. Either way, it acts as a convenient and in-depth manipulation check.

The granularity afforded by this deeper level of analysis allows us to be more exact about the mechanisms behind any effect detected in the larger population. For example, in F. Vallée-Tourangeau et al. (2020), participants were invited to solve the triangle of coins problem in a low interactivity condition and a high interactivity condition. Crucially, they were only allowed one guess[50] at the solution. While the high interactivity yielded more successful trials those trials were significantly slower. A qualitative evaluation of video material showed that in the high interactivity condition participants would solve the problem but because the interface afforded them an easy opportunity to check their answer, they would take advantage of this, build the solution a second time and so increase their latency. So, while the condition was the cause of the difference in latency, the mechanisms through which it caused that difference were only revealed by case study analysis. The material traces of the physically realised actions involved in problem-solving allowed the thought process to be traced. A simple measure of latency would have assumed slower cognitive processes.

This level of analysis is useful to unpick the effect of the experimental manipulation was which in turn allows us to test some of the models which are being proposed. I have cast problem-solving as a process and so it becomes important to map this process with a high level of granularity. This behaviour observation will allow us to draw

---

[50] In Experiment Two, in Experiment One they were allowed unlimited guesses which I discuss at greater length in Chapter Six.

conclusions which could be tested further in experimental situations. Leaving aside theoretical concerns for now, it is unclear how novel explanatory mechanisms for mental processes could be generated without the observation of behaviour. The benefit of placing participants in first order problem-solving environments is not just that the potential benefits of this type of environment can be assessed but also because the material traces of problem-solving can be exploited to guide our understanding of how problem-solving progresses. This return to data and to behaviour in action reflects a pragmatist perspective rooted in observable action.

Like Steffensen's Cognitive Event Analysis (Steffensen et al., 2016), this level of analysis does not require naturalistic material. Indeed, in these cases the analysis is bound to a single problem-solving moment generated by an experimental situation. Thus, while the analysis deals with a small number of cases, the function of the analysis in this case is more reductive than typical case studies: it is temporally and artificially bounded around the cognitive tasks. It is furthermore part of a pre-planned mixed analysis plan, its purpose is to elucidate conclusions drawn from the experimental manipulations and the statistical analysis of data sets. In addition, a case study analysis can allow us to avoid attributing behaviours to an average participant who does not exist (as in Cushen & Wiley, 2012)

While it can be understood on its own, it is not designed to be read in such a way and the conclusions drawn and the levels of analysis required are driven by a requirement to understand and enrich the quantitative data. Its primary function is exploratory and descriptive rather than explanatory (Yin, 2014) because the causal mechanisms implicit in the experimental form provide the hypothesised explanatory mechanisms. In many ways this level of analysis functions to replace the discussion section of a typical quantitative only research paper which is replete with – 'it is plausible that', or 'informal observations suggest that' which are offered as explanatory mechanisms beyond the aggregated means. Furthermore, such qualitative work gives empirical support to principled hypothesis formation (Steffensen et al., 2016) and stretch out the boundaries of a theoretically driven research programme. It is not intended to replace quantitative and aggregated results but rather enhance the understanding of the mechanisms through which the effects of the experimental manipulation are realised. Bennett and Elman (2006) contrast these two aims thus: quantitative research takes an effect of causes approach by manipulating the cause and measuring the effects whereas qualitative research takes a causes of effects approach.

The Nature of the Qualitative Enquiry

What defines qualitative research is often a negation: Quantitative research is numbers and qualitative research is what is left, often text based (Ketokivi & Choi, 2014). The analysis I propose here takes a highly granular and qualitative approach looking at one or more critical cases. It looks more broadly, although in detail, at bounded cognitive events where the population of principal inference is the experimental population. It is not intended to replace quantitative and aggregated results but rather enhance the understanding of the mechanisms through which the effects of the experimental manipulation are realised. Bennett and Elman (2006) contrast these two aims thus: quantitative research takes an effect of causes approach by manipulating the cause and measuring the effects whereas qualitative research takes a causes of effects approach. I argue that kinenoetic analysis as described here can strengthen both understanding of cognition but also how thought can be traced beyond text-based approaches which necessarily centre the experience of the human agent: Qualitative work addresses process through the understanding of the participant (Lahlou, 2011). However, this focus inwards, to an inner phenomenological but this is not the aim here and indeed the qualitative nature of the data should be dissociated from a qualitative epistemological position. There is not a necessary connection (see Chapter Eight for an in-depth discussion).

The Methodological Aims

Systemic approaches to higher order cognitive processes (F. Vallée-Tourangeau & Vallée-Tourangeau, 2020; G. Vallée-Tourangeau & Vallée-Tourangeau, 2014) are designed to examine cognition as an emergent property of a system of people and things. This research programme has as its explicit aim to move the focus of cognition from the brain to a coupled brain/world system. However, the first wave of this research was still firmly experimental and drew boundaries on what constitutes a part of the system—boundaries which were recognised as artificial but boundaries, nonetheless. The boundaries were also temporal and carved problem-solving time into two states: a starting state (one of ignorance) and an end one which is judged against the normative values (correct or incorrect). However, despite these boundaries, a systemic research programme opens the system in a hope to provide a richer understanding of cognition. Experiments in this tradition place the problem solver in a first order problem-solving environment, that is the problem-solver is embedded in a materially rich environment and allows the manipulation of these objects. The second wave of this research programme is still in infancy although the programme of this methodological interactivism have been sketched out in F. Vallée-Tourangeau & Vallée-Tourangeau (2020) and this thesis aims to build on the methodology proposed therein.

The research reported in this thesis has three main aims. First, to establish whether people do naturally recruit the world around them to structure their problem-solving and under what circumstances they do this. Second, to move the focus from profiling performance to profiling the object-thought mutualities as they arise and to assess the additional explanatory factors that arise.

What Do People Actually Do?

F. Vallée-Tourangeau and March (2019, p.6) suggest that "it does not require much ethnographic effort to notice people think with their hands, their body, and wide range of disparate artifacts" but do not give evidence of this beyond the expansive gesture to 'what we all know', a folk psychology of how we interact with the world around us. A similar expansive gesture is made by other authors such as Wilson and Clark (2009). This is problematic; while it is clear to see from anecdote and from personal experience that people recruit[51] the world to structure their thinking, anecdotal evidence and evidence from personal experience could suggest that this is not always the case. Take for example the experience of closing one's eyes and shutting out the world to fully concentrate and retrieve an idea. If the proponents of first order problem-solving appeal only to common sense as a reason to embed problems in the messy real world then they are likely to be challenged. As we shall see in Chapter Six, such intuitions can sometimes be problematic.

A traditional experiment in systemic thinking will contrast a high and a low interactivity problem-solving environment. A low interactivity environment is similar in character to the second order problem-solving environment described by F. Vallée-Tourangeau and March (2019). The participant is immobilised to a greater or lesser extent and problem-solving performance is necessarily a reflection of internal processes. A high interactivity environment is more of a first order problem-solving situation where

[51] The agent centric notion of this verb revealing a commitment to a first wave of extended mind theories.

participants are invited to interact with and manipulate external representations such as numbered or lettered tokens or physical models of the problem. Performance is generally calculated as mean differences in success rates across the two conditions. There is little assessment of how participants actually behave in the high interactivity condition, the assumption is that all participants adhere equally to the conditions. The difficulty with this is that there is likely to be a continuum of behaviour, with some participants in the low interactivity conditions talking to themselves, moving bodily[52] and using the full extent of moves allowed to them depending on the constraints (e.g., a paper and pencil or gesturing) whereas it is equally possible that in the low interactivity condition participants choose not to move at all. Indeed, Maglio et al., (1999) report that roughly one third of the participants did not, in fact, use their hands or used their hands very briefly despite being at complete liberty to do and despite this being the key experimental manipulation. That a significant minority of the sample did not consider it worth using their hands to structure their thoughts, requires us to consider to what extent the participants in these conditions could be said to be using interactivity – rather the condition might be more aptly renamed ‘potential for interactivity.

Therefore, given that it is an underlying assumption of research in distributed cognition that people will naturally recruit their environment and that this is a justification for this research programme, it needs to be empirically established. The easiest way to do this is to record participants’ behaviour and catalogue the time spent adhering to the experimental manipulations. It will also allow the central theoretical tenet that people do use the world around them when solving problems to be tested. It has further theoretical

[52] In the experiment reported in Ross et al. (2019), children were observed who were required to sit on their hands moving their whole bodies to aid their thinking.

implications for understanding the nature of the coupling of nature and the world. There is an implied central locus in the idea of an extension as a choice which points to a loosely connected joint. This weakness can be remediated by establishing that we do indeed regularly recruit the world when we are thinking and moving the focus to considering that nature of the coupling that ensues and the changes to processes rather than simply measuring performance. Chapter Four will establish the importance of monitoring engagement, Chapter Five will assess the nature of that engagement, while Chapter Six will examine other ways we bring in the world in thought.

Additional Explanatory Factors

A focus on process rather than performance alongside an increase in the granularity of the observations, allows the development of a bottom-up theory of problem-solving and thinking. That is a theory which is based in observation rather than theory. It is important to maintain epistemic humility. The same messy decision making which is erased from experimental reports outlined above is still evident. Take Köhler's observational studies of ape problem-solving behaviour. The observation of Sultan here informed Gestaltist views on insightful problem-solving but as F. Vallée-Tourangeau et al., (2020) argue a focus on other the apes (namely Koko!) would have revealed very different behaviours and perhaps inspired a different line of research. The proposal for methodological change put forward in this thesis does not lay a claim to purity or superiority but a adopts a pragmatist focus on behaviour, on what people *do*. Thus, by using observational and idiographic methods, the work presented in this thesis will explore some uncharted territory in our understanding of how thinking happens. We can stop "over intellectualising the intellect" (Noë, 2012, p.119) and assuming that all thought is the product of effort.

By viewing experiments not as measures in support or not of hypotheses but as tools to observe the formation of soft assembled cognitive systems under controlled circumstances, we open up the research field to other explanatory factors. Take for example the role of accidents: When participants are provided with cognitive artefacts to scaffold their thinking either as part of intentional focus as in research in interactivity or as an unconsidered artefact of the environmental set up, the idea of chance is often mentioned. Fioratou and Cowley (2009), for example describe a version of the cheap necklace problem and suggest that 6 of the 21 solvers (almost a third) solved the problem through the exploitation of an accident. This same observation occurs in Chuderski et al., (2020, p.18) who suggest that "in the matchstick algebra problem, it is arguably easier to arrive at the solution by accident or trial and error, for instance by realizing as a result of a random movement of a stick that it could act as a negative sign." In one of the few qualitative approaches to insight problem-solving, Steffensen et al. (2016)'s finely grained analysis of problem-solving demonstrates that the solution hangs on an accidental moment.

The role of accident here may be an important explanatory factor but if it is recognised it is subsumed into existing models. So, for example, we can see Fleck and Weisberg (2013) describing so-called data driven restructuring as generated by the environment: 'Data-driven restructuring included instances when the individual changed his or her representation of the problem in response to something he or she saw from the physical configuration of the problem […]. Observations occurred as the participant was attempting to construct or implement another heuristic-based solution.' (Fleck & Weisberg, 2013 p.452). However, observations such as these, that there are times when the environment yields a solution to the problem, are under investigated precisely because the theoretical and methodological allegiances mean that it is simply missed. If the aim of problem-solving research is to uncover the internal algorithms and processing that leads to a solution, then the role of accidents can shed little light on this. If we are interested in problem-solving as it unfolds, not as a reflection of hypothesised and ontologically unstable states then it behoves us to take all possible explanatory factors seriously.

Conclusion

The studies reported in this thesis are all experiments. They are the results of pushing and prodding nature to see what will unfold and to provoke certain behaviours at a convenient time and place. Chapters Four, Five and Six are conducted within an experimental paradigm consistent with previous research on interactivity and systemic thinking (see for example Fioratou & Cowley, 2009; Henok et al., 2018; Maglio et al., 1999; Vallée-Tourangeau, 2013). However, the focus of the research presented here is significantly different from previous research into interactivity and systemic thinking because it shifts the focus from performance to behaviours. In so doing, I will demonstrate that such research programme is ill-equipped for dealing with the complexities of research in an open system.

It is tempting to think that extending the cognitive system must necessarily strengthen it. This is often the case where the system acts as a simple offload but it is not a necessary quality of the extended mind (Wilson, 2014) and indeed as the research programme into materially and socially extended cognition matures, it is becoming clear that there is not a straightforward additive relationship. Whereas interactivity confers a boost to tasks requiring working memory by reifying thoughts and expanding the mental workspace, a task which requires the problem solver to reject incremental steps may not incur the same benefit from this workspace extension. For example, many insight problems have an inviting initial presentation which interactivity would make easier to follow but which is ultimately unhelpful such as those which appear to require arithmetic but instead require the participant to disregard arithmetic (such as I will demonstrate in Chapter Six). Indeed, the reification of unhelpful initial presentation states or procedural steps may make it harder to disregard these and lead to a reduction in performance. This is not to suggest that interactivity is to be disregarded as component of cognition, rather, that a simple linear formula is unhelpful in characterising the cognitive performance emerging from a complex system of internal and external parts. Indeed, that interactivity can sometimes have a detrimental impact on performance but still forms part of the system is perhaps a stronger argument for its constitutive role.

Chapter Four : Interactivity in a Word Production Task

The central argument of this thesis is not that cognition cannot happen without environmental engagement but that experiments in cognitive psychology should (i) design task environments which allow engagement with movable materials to better track situated cognition and (ii) use these material traces as a research tool to support inferences about problem-solving process. Thus, interactivity becomes not only a theoretical position but also a methodological tool to track thought in the world. This kinenoetic analysis (KA) allows us to use the movement of objects in the world as a way of tracing environmentally generated cognition.

Once the premise that environmentally situated cognition is constituted by aspects of that environment is accepted, the ensuing premise becomes self-evident: Even cognition which unfolds in a vacuum will reflect the nature of that vacuum rather than a set of universal principles. The cognitive processes elicited in this vacuum may yield no information about the processes used outside of it (see Chapter Six for a longer discussion). Therefore, task environments need to be designed to take this into account and novel performance measures should be designed.

The empirical investigation of the argument begins with a conceptual replication of Kirsh (2014). This study introduced the role of chance in to the classic word production task (previously used by Fleming & Maglio, 2015; Maglio et al., 1999; Vallée-Tourangeau & Wrightman, 2010; Webb & Vallée-Tourangeau, 2009) and so was selected as a particularly interesting way of examining previously unconsidered parameters in problem-solving. This chapter will describe this conceptual replication which moved the task environment from one with digitally represented letter sets to a task environment in which the letters were represented on actual physical tokens. The replication also extended

Kirsh's study by undertaking a detailed analysis of the movements of the tokens to increase the understanding of the mechanisms underlying performance on this task in line with the methods of KA.

The Word Production Task

If problem-solving is best understood as an activity that takes shape in a dynamic meshwork of resources and processes, configured from internal mental resources, embodied actions and environmental affordances then the challenge posed by this meshwork of people and things is to design a research platform that offers a window onto relatively open-ended problem-solving that is nonetheless amenable to detailed quantitative and qualitative analyses. Such a platform should help us determine the role of an agent's internal resources or cognitive capacities and her actions as cued by the affordances offered by a dynamic and malleable external environment.

One such platform is Maglio et al.'s (1999) word production task. In this task the game of Scrabble™ has been modified to assess whether manipulating the letter array supports participants in generating new words, or word production. Scrabble ™ involves using 7 letters to create words. In the game, the letters are scored in terms of frequency and also by their position on the board. The task has been adapted by cognitive psychologists to assess the circumstances which help word production (Fleming & Maglio, 2015; Kirsh, 2014; Maglio et al., 1999; F. Vallée-Tourangeau & Wrightman, 2010; Webb & Vallée-Tourangeau, 2009). In this modified task, participants are given 7 letter tiles and asked to use those letters to generate words which they announce to the experimenter for a short period of time, commonly 3 or 5 minutes. It differs from Scrabble ™ along three dimensions (a) the letters remain the same throughout (b) the letters are unscored (although the words produced can be scored subsequently, the participant is not motivated to aim for high scoring words or to use difficult letters) (c) the words produced are not necessarily contingent on previous words, although they can follow on they do not need to be valid.

While closely related to anagram tasks, the word production task also differs from these because there is not one right answer and so the skills involved are qualitatively different. Smaller words can be produced and the rate of production rather than the proximity to a right answer is measured. This makes it an open problem with more flexibility and yet one which invites a higher granularity of analysis. Success in the task therefore requires a complex set of skills from searching an existing vocabulary space to the recognition of productive words stems and so relies on both working and long-term memory. As the outcome measure is the number of words then the rate and fluency of processing is also important.

Additionally, participants' performance can easily be observed in two distinct task environments: In a low interactivity environment, where tile movement is constrained or forbidden, performance reflects mental resources alone, while in a high interactivity environment, participants can move the tiles as they see fit in supporting word production. As outlined more fully in Chapter One, in the low interactivity condition, the problem solver is decoupled from her immediate environment: She is invited to solve a problem without using her hands to support thinking either through gesture or rearranging the physical elements that configure a model of the problem (such task environments are often the default procedure employed in problem-solving research). In other words, problem-solving proceeds from mental simulations of possible solutions. In contrast, a high interactivity task environment places no such constraints on her: Participants are presented with physical elements of the problem that can be manipulated to arrive at a solution. In such environments, proto solutions unveil new action affordances and guide attention in ways that are simply not possible in low interactivity conditions. The reification of the problem scaffolds working memory resources and the engagement with a movable environment increases the likelihood of serendipitous discovery. Finally, the use of an interactive task environment also facilitates the use of KA to document the problem-solving process (see Chapter Three). That is, the movements of the tiles constitute visible indications of thought and the lettered interface allows the detailed mapping of strategies and reveals cognitive movements and actions.

Kirsh (2014) used this task to test more explicitly the role of luck in word production. Alongside a low and high interactivity task environment, he introduced a shuffle condition to isolate the element of chance. The participants were asked to call out as many words as they could in three minutes from a set of seven letters presented on a computer screen. They were asked to do so in three conditions: interactive (when the letters could be rearranged using a mouse), static (the letters could not be moved) and shuffle (one click shuffled the letters randomly). Kirsh found that the shuffle condition produced a significantly higher number of words than either the interactive or the static condition[53]. This element of luck is particularly important when considering underexamined additional explanatory mechanisms through which a transactional relationship between artifacts and problem solver can moderate outcomes.

---

[53] Participants produced the most words in the shuffle condition ($M = 18.88$, $SD = 4.14$), followed by the interactive condition ($M = 17.67$, $SD = 4.31$) and then the static condition ($M = 16.56$, $SD = 5.35$); the main effect of condition was significant, $F(2,50) = 11.28$, $p < .001$. Post-hoc t tests revealed that word production performance was significantly higher in the shuffle condition than in the interactive condition, $t = 2.49$, $p = .02$, and significantly higher in the interactive than in the static condition, $t = 2.20$, $p = .038$ (Kirsh, 2014, p. 17).

Performance Moderators in the Word Production Task

There are clear theoretical reasons to suppose that interactivity would benefit solvers in a word production task. By extending the mental workspace outside of the head, the internal letter representations are reified and are easily manipulated freeing up working memory and scaffolding participants' internal resources (Gavurin, 1967; F. Vallée-Tourangeau & Wrightman, 2010; Webb & Vallée-Tourangeau, 2009). Furthermore, interactivity allows the solvers to move with less effort through the problem space and even to jump to new places with, at times, unplanned moves (Maglio et al., 1999). Thus, the tiles may either be recruited strategically or, more serendipitously, non-strategic moves may yield lucky combinations of letters.

However, the data supporting a scaffolding effect of a high interactivity environment on a word production task are less clear that might be imagined. The only experiment that demonstrates an unequivocal benefit is reported in Fleming and Maglio (2015) where interactivity not only led to an increase in word production but also to rarer words being produced. While Maglio et al. (1999) documented a small overall benefit for interactivity, when this was broken down into the two different letter sets used, interactivity led participants to produce more words with one, harder, letter set but fewer words with another. With the easier letter set, participants are more capable of generating words without help so the added cost of manipulating tiles may actually slow down word generation. Interactivity is only worth pursuing when the costs do not outweigh the benefits (Kirsh, 2009; Rowlands, 2010). In a balance of costs where there is incentive to produce a large number of words over better quality words (three 2-letter words are worth more than one 7-letter word) then there is not enough of a benefit conferred by the time spent rearranging the tiles to make it worthwhile. If, however, the word set were harder or the incentive was for longer words then, perhaps that interactivity becomes more important. Rowlands (2010) proposes the crucial caveat to cognitive offloading: It is not worth the cognitive cost of offloading if the work involve in offloading is greater than the efficiency gains. If a word can be instantly generated through merely “offline” cognition the time and cognitive cost of moving letters may even lead to a lower rate of production in a high interactivity condition. A review of the existing literature suggests three main moderators of performance in this task: internal dispositions, enacted behaviour and environmental affordances.

Individual Differences in Verbal Fluency.

Where individual differences have been profiled, the internal resources of the participants moderate the benefits of interactivity. F. Vallée-Tourangeau and Wrightman (2010) found that there was a statistically significant benefit for participants categorised in a low verbal fluency group while the benefit for those in the high verbal fluency group was negligible. This mixed story is echoed by Webb and Vallée-Tourangeau (2009) who looked at the benefits of interactivity in children with and without developmental dyslexia. The control group produced marginally more words with their hands when presented with the harder set of letters condition and the difference with the easy set was indistinguishable. The children with developmental dyslexia (who displayed lower working memory score and lower verbal fluency) benefitted from interactivity with the easy letter set but there was no difference with the harder set. This indicates that interactivity can have a benefit, but that benefit is not universally enacted and is tied to variations in individual difference profiles and letter set difficulty which further points to a more complex relationship than the straightforward additive one normally suggested.

Individual Differences in Behaviour.

In a footnote, Maglio et al. (1999, p. 330) report that roughly one third of the participants did not, in fact, use their hands or used their hands only very briefly despite being at complete liberty to do so and despite this being the key experimental manipulation. That a significant minority of the sample did not consider it worth using their hands to structure their thoughts, requires us to consider to what extent the participants in these conditions could be said to be using interactivity – rather the condition might be more aptly renamed ‘potential for interactivity’. In various experiments investigating the role of interactivity in problem-solving (e.g., F. Vallée-Tourangeau et al., 2016), the low interactivity condition is invariably tightly controlled, and participants’ movements are constrained with them often being requested to lay their hands flat on the work surface. However, there are few controls and rarely any consideration of the manner in which participants recruit resources in a task environment labelled as high interactivity. Only Fleming and Maglio (2015) took closer look at the behaviour of participants in a high interactivity condition in this task. In contrast to Maglio et al. (1999), they suggest that all their participants moved the tiles. However, as the focus of their paper was strategy selection rather than the time spent interacting, the detailed analysis required restricting their coding to the behaviour of only 8 participants in the final block with a specific aim of looking for and coding word production strategies. If we are to profile the whole system (G. Vallée-Tourangeau & Vallée-Tourangeau, 2017) then the level of interaction becomes important each time the participants encounter the tiles as a measure in itself rather than solely as an indication of strategy, especially if the different levels of interactivity designed in these environments do not result in differences in participants’ behaviour.

It is further unclear how much participants' behaviour differs as a function of their individual differences. This was addressed by G. Vallée-Tourangeau et al. (2015) in research into whether interactivity can scaffold Bayesian reasoning. They found that successful reasoners were more likely to engage in presentation change whereas unsuccessful reasoners were more likely to engage in marking (that is, simply touching the material, rather than rearranging it). A presentation change involved actively changing the layout rather than just nudging or moving the cards. So, it seems likely that "interactivity" itself is not a panacea, rather there needs to be a meaningful interaction with the representation.

It is plausible that those who do not need the help of the tiles recruit them less. Research on expert Tetris players suggests an inverted U shape relationship of action and expertise with complementary actions decreasing as expertise increases (Destefano et al., 2011) Kirsh (2014) also reports a significantly different strategy between the best and the worst performing participants reflected in the number of shuffles they elected to perform. Excess shuffling represented a disadvantage either because the participants who shuffled more were less able to produce words and so chose to shuffle more frequently to break their impasse or because shuffling too much was itself a poor strategy. This indicates that the relationship between condition and performance is also mediated by the behaviour within the conditions. It is not unreasonable to expect a similar relationship in this task.

Environmental Affordances.

Across the different uses of the word production task there have been different task environments in the high interactivity condition which have not been fully taken into consideration. The focus of research has been how being able to manipulate the array is the primary benefit, the nature of the manipulation and the affordances of different interfaces have been of less interest. This can be seen by the wide variety of interfaces used: Across only five studies we see three different ways of interacting with the letter displays which are not commented on by any of the authors (see Table 4.1).

Table 4.1

The Different Material Provided in Each of The Word Production Tasks

Study	Nature of the Materials Provided
Maglio et al. (1999)	Tiles from the Scrabble game
Webb & Vallée-Tourangeau (2009)	Not stated.
Vallée-Tourangeau & Wrightman (2010)	Not stated although "tiles" is mentioned
Fleming & Maglio (2015)	iPad (using touch and drag).
Kirsh (2014)	PC (using point and click or one click shuffle).

As there were no reported constraints in the high interactivity condition in Kirsh (2014), this condition actually affords the widest range of potential strategies: It is theoretically possible to shuffle, move the tiles at will or simply leave a static array. In practice, it seems unlikely that participants could have fully used the range of available possibilities of the high interactivity version. Indeed, if we consider the behaviour in the shuffle condition—the best performing third shuffled once every 3.7 seconds, the worst performing third once every 1.9 seconds— the ease with which the shuffle could be executed using the shuffle button is clear and demonstrates the short time and low cognitive cost of generating environmental hints in the condition. Given the more cumbersome mouse interface, it would be impossible to mimic this strategy with the high interactivity version as speedily or effectively. However, this is a function of the affordances of the digital interface which invited participants to interact with the tiles by either pushing a button in the shuffle condition or selecting with a mouse and dragging the tiles into place. It is my hypothesis that in Kirsh (2014) the low cost of shuffling the tiles with one click on a computer compared to the relatively high cost of moving tiles with a mouse, meant that shuffling functioned as an epistemic action (Kirsh & Maglio, 1994) more closely resembling the actions of a Tetris player who chooses a tetromino drop location based on what she sees *after* multiple physical rotations rather than a true test of luck. Indeed, Kirsh acknowledges this: "the cost in time and mental effort must be sufficiently low that it pays to keep fishing for hints" (Kirsh, 2014, p. 19).

Just as the skills of the problem solver are important in the interaction between person and material world, so too are the affordances for action offered by the artefacts. Taking the cognitive ecology of this task seriously, requires taking the affordances of the external environment seriously. Rather than making the implicit assumption that the problem solver imposes her will on an inert and indifferent environment, it seems plausible that the nature of the artefacts selected will determine to some extent the actions undertaken (Steffensen et al., 2016). Indeed Kirsh (2014, p. 8) notes: "we think and perform differently in different media." As different concrete task environments afford different actions and strategies (Guthrie & Vallée-Tourangeau, 2015), it is likely that there is a reciprocal relationship. It is becoming increasingly clear that research in interactivity requires a more nuanced analysis of the nature and the process of coupling between mind and world (see Chapter One for more on this second wave of interactivity research and Chapter Seven for a closer examination of the relationship between material and agent through the lens of Material Engagement Theory).

An Interactive View

The presentation of these three aspects—differences in underlying cognitive capacities, differences in behaviours and differences in environmental conditions—implies a static tripartite approach. Rather, in practice, the differences between the three will be difficult to disentangle and there is likely to be an interactive and reciprocal relationship. There is an interaction between abilities and affordances (Chemero, 2011). A chair, for example, affords sitting for most adults but for a child learning to walk affords support to get from one place to another. Therefore, to discuss the affordances of a problem set up without considering the affordances *for whom* is immediately problematic. Similarly, the direction of the link between behaviour and problem solution is sometimes unclear. In the study described by G. Vallée-Tourangeau et al. (2015), it is hard to judge if it is the type of interaction which drives a successful problem solution or to determine whether those who are likely to solve the problem employ a certain strategy. Performance on this task is likely to be predicted by a complex mix of these three things. Therefore, it is also important to track whether there are individual differences which predict the level of engagement with the outside world.

The Benefits of the Shuffle Condition.

For Maglio et al. (1999) the benefit of high interactivity in this task was in no small part due to the introduction of possible randomness that seeds intelligent behaviour. A shuffle condition isolates this moment of chance. In this task, a beneficial strategy is to use the same root and change a single letter. In this way a participant may identify the root *-ate* in the word set and produce the words d-*ate*, f-*ate*, g-*ate*, h-*ate*, l-*ate* and so on depending on the other letters available in the set of 7 – a switch between state spaces (Maglio et al., 1999). This is an efficient strategy until the words run out and then another root will need to be identified necessitating a larger shift in the problem space. If this randomness is an additional basis for the benefits of interactivity by introducing participants to elements of the state space they would not have mentally simulated or faster than they could have mentally simulated, then this should be the main advantage of the shuffle condition.

Indeed, the idea that problem-solving may be augmented by unplanned changes in the environment seems to occur regularly when problem solvers are given an environment that allows for such accidents. For example, Ormerod et al. (2002) invited participants to solve the 8-coin problem using hexagonal coin tokens and write of "serendipitously encounter[ing] an external object or event" (p.797). In the same way, when Fleck and Weisberg (2013) presented participants working on the triangle of coins with actual coins to build a model of the solution, the researchers noted that some solution hints could be data driven, that is offered by changes in the physical configuration of the problem (that might or might not been premeditated). This is perhaps most clearly discussed by Seifert et al. (1994) whose opportunistic assimilation theory assesses the characteristics of the 'prepared mind'. However, a systematic examination of this contingent and emergent phenomenon which assesses the problem solver as part of a coupled system has yet to be undertaken (see Chapter 3 for a more detailed examination of the role of accidents in problem-solving as a yet unexplored mechanism).

Serendipity

Kirsh is explicit that he views chance mainly as the provision of hints. Despite the high rate of shuffles facilitated by his task environment, he rejects the idea that this constitutes a variation on the idea of blind variation first put forward by Campbell (1960) and later explored in more detail by Simonton (see for example Simonton, 2010). Kirsh suggests that chance is important as a way of introducing diversity in hints which, crucially, have to be exploited by the problem solver. In this way, he is moving from a non-agentic vision of luck (Coffman, 2009; Griffith, 2010) to one which requires a systemic approach. This can be more properly understood as serendipity. Chance is inert until it is exploited by an agent. No matter how much luck is introduced into an environment, it only becomes useful when it is used to produce an outcome. This interaction of environmentally induced chance and human agency is known as serendipity.

Serendipity has been described as being “at the intersection of chance and wisdom” (Copeland, 2019, p. 2385). The serendipitous process emerges from an interplay of external and internal factors exactly as posited by Kirsh (2014, p. 14): “the chance event must be interpreted and worked with.” It is not enough for the environment to create lucky situations: an individual must take advantage of those situations. Equally, an individual’s internal resources are not the sole actant in cognitive activity – the opportunities thrown up the environment are equally important. The central concept of serendipity suggests that the benefits are not indiscriminate and instead rely on the dialogue between person and environment.It is further important to note in this regard, that Kirsh reports a significantly different strategy between the best and the worst performing participants reflected by the number of shuffles. Excess shuffling represented a disadvantage but the mechanisms behind that are not clear: Either the participants who shuffled more were less able to produce words and so shuffled more as compensation or because shuffling was in itself a poor strategy. This indicates that the relationship between condition and performance is also mediated by strategy and potentially individual differences The scrabble task is a particularly interesting interface to explore the interaction of chance, movable artifacts and internal dispositions because there is empirical evidence to support the importance of all three aspects to performance. An explanation of the mixed data to date could be that performance on this task is an emergent property of these three factors. Moreover, because it requires the generation of many words, it consists of multiple, small problem-solving opportunities allowing plenty of opportunities for serendipitous moments to occur. Therefore, a profile of individual differences and an attention to strategy are essential to fully understand the cognitive processes in this task.

The Current Experiment

The current experiment adapted Kirsh's (2014) procedure and transferred it to an environment which would be more likely to increase the time and cognitive costs of the shuffle condition to assess whether the role of luck extends across environmental presentations. At the same time, moving the procedure to a physical rather than a digital environment afforded a greater range of movements in the high interactivity environment. It was predicted that this would change the behaviour of the participants and elicit more moves in the high interactivity condition and fewer shuffles in the shuffle condition reversing the performance outcomes reported in Kirsh (2014). It was, therefore, hypothesised that the high interactivity condition would yield higher performance, followed by the shuffle condition and then the low interactivity condition.

The participants were also profiled across the individual differences thought to be particularly relevant to this task in light of the evidence from Webb and F. Vallée-Tourangeau (2009) and F. Vallée-Tourangeau and Wrightman (2010). Additionally, filming participants, enabled a more granular analysis of the strategies employed which allowed a clearer understanding of the complexities that drove performance in the high interactivity and the shuffle conditions and develop data-grounded exploratory hypotheses. It further functioned as a manipulation check to assess the numbers of participants in the high interactivity condition who chose to move the letter tiles and the manner in which they chose to do so. It was hypothesised that while moving the tiles in the high interactivity condition would be beneficial, it would show an inverted U relationship as already noted in Destefano et al. (2011) such that too little or too much movement would be detrimental to performance. Analysis of the relationship between individual differences and the amount of time interacting could indicate whether this was because the strategy itself was not beneficial at the extremes or whether the strategy reflected different ability levels in the task.

The current study had three parts: a norming study to establish consistency across the experimental stimuli, a technical pilot and the main study. The reports and data for the norming study and the technical pilot can be found in **Appendix A** and the main study is reported here.

Method

Participants

Forty-two psychology students (*F*=33; M_{age} = 25.4, *SD* =7.2) at Kingston University were invited to participate in return for course credits. Two declined to be filmed and their data was removed from analysis. This left 40 participants (*F*= 32) aged between 19 and 47 (*M* = 25.65, *SD* 7.22). All had either had English as a first language or considered themselves fluent speakers of English.

Materials and Measures

Letter Sets

Three sets of letters were selected after an online norming task administered to a group of online participants (N=134) via the Qualtrics platform to establish which were most similar in terms of words produced and their relative frequency in a naturally occurring corpus of utterances (Zipf score)[54].. Six letter sets were selected (for more details see **Appendix A**) Each participant was asked to find as many words as they could from the set of seven letters in five minutes following the same procedure as in the proposed study only in an online form. Each participant was invited to consider two of the remaining six letter sets. The letter sets and the order were randomly selected by the Qualtrics software. The words they produced were analysed for word count, number of letters and word frequency Those three which were most similar in characteristics were selected as stimuli for the task and are presented in Table 4.2.

Table 4.2

The Average Number of Words, The Average Letter Count and The Average Frequency of Words Produced in Five Minutes For Each of the Selected Letter Sets

	N	Number of Words	Letter Count	Zipf Score
COTFAED	41	20.56	3.68	3.95
NDRBEOE	39	20.33	3.81	4.12
TVAERWI	38	19.95	3.78	4.14

Note: The varying numbers of participants are because of the randomising in Qualtrics

Outcome Measures

The participants' performance was measured in terms of (a) the total number of words produced, (b) the total word count and (c) the frequency of the produced words as a Zipf score (see footnote 42).

Measures of Individual Differences

Verbal fluency. Verbal fluency was measured with an adapted Thurstone Task

[54] The word frequency scores in this thesis were based on the Zipf scores presented in the SUBTLEX-UK database (van Heuven et al., 2014). This database takes word frequencies on the basis of subtitles of British television speakers and coverts these to a 7 point logarithmic scale. Note that the lower the Zipf Score the less frequently the words occur. We would therefore judge a low Zipf score as reflecting a deeper search space.

(Thurstone, 1938). The test invites participants to write as many words as possible beginning with the letter "S" for five minutes, and then as many words as they could beginning with "C" for four minutes. The participant's score was the total number of words produced in the 9-minute period.

Anagram Skills. The cognitive skills required by the word production task are not just fluency of production but also the generation of novel words from a set of letters, essentially creating a series of different anagrams. It is therefore perhaps unsurprising that there are established correlations between Scrabble expertise (on a national rating scale) and anagram skills: expert Scrabble players outperform average players and both outperform novices (Tuffiash et al., 2007). The correlations between Scrabble skills and anagram skills suggest that this individual difference may moderate performance on the word production task, indeed, Friedlander and Fine (2016) also reported that cryptic crossword solvers have naturally good anagramming skills which has a strong parallel with the current study given that these solvers also tend to deliberately restructure letters. Twelve anagrams were selected from the set used in Webb et al. (2018) which were in turn drawn from Novick and Sherman (2003). Each anagram was solvable within one-, two- or three- letter moves with two-letter moves being the most common. Twelve were selected to minimize the potential for participant fatigue during the study. The anagrams selected were administered as part of a larger set of 35 anagrams to 544 participants by Webb et al. and the anagrams selected yielded an average solution rate of 73%. The anagrams, their solutions and the percentage solution rate are shown in Table 4.3.

Table 4.3

Anagrams Selected for the Task, the Correct Solution and the Overall Solution Rate as Reported in Webb et al. (2018)

Anagram as presented	Correct Solution	Solution Rate
Iasdy	daisy	79%
Nadts	stand	81%
Dnsuo	sound	85%
Injto	joint	83%
Lcrue	cruel	95%
Nrtai	train	82%
Ichem	chime	34%
Opchu	pouch	78%
Elbaz	blaze	42%
Leogv	glove	69%

The anagrams were presented on Qualtrics. The participants were given 30 seconds to come up with a solution to each anagram. They were not allowed to use any external aids such as pen and paper. They were not automatically given the answers to the anagrams but were given the answers at the end of the session if they requested them. Only one participant requested them

Extraversion and Openness to Experience. There is limited research on which personality traits correlate most strongly with this propensity to take advantage of luck in the environment, but it has been suggested that extraversion is a salient personality dimension (McCay-Peet et al., 2015). I assessed participants along this dimension using the extraversion elements from a revised version of the same scale (the HEXACO-PI-R; Lee & Ashton, 2018). I also profiled the participants on the openness to experience elements of the same scale given the suggestion that this personality trait is related to the processing of irrelevant information (Agnoli et al., 2015; Friedlander & Fine, 2018).

Design and Procedure

The experiment used a repeated measures design with the order of the three experimental conditions counterbalanced across participants. The three conditions were high interactivity (high), low interactivity (low), and shuffle. The sets of letters (Table 4.2) also counterbalanced across conditions. The participants were invited to call out as many words as they could from a set of 7 letters. The instructions for the high interactivity conditions read as follows:

> Look at the 7 letters in front of you. In a five-minute period, I would like to you to make as many words using these 7 letters as you think possible. Words need to be at least 3 letters long and I encourage you to make full use of the range of letters. To help you do so, you can touch and re-arrange the letter tiles in front of you however you wish. You can re-arrange the tiles in any manner you wish. When you think of a word, say it first, then spell it, I will write it down. For example, if you think of the word 'dog', say it first 'dog' and then spell it 'DOG'. Proper names and acronyms (e.g., IBM) are not acceptable.

In the low interactivity condition participants were asked to not interact with the tiles in any way or to use their hands to gesture or to point. If they did not adhere to these instructions, the experimenter reminded them and requested they hold their hands in their lap. To avoid the low interactivity condition itself carrying a cognitive cost, participants were not asked to hold their hands firmly in place on the table. It was considered that the chance of a cognitive load generated by holding hands on the table would be greater than the chance that micromovements would affect their performance.

In the shuffle condition, participants were told they could shuffle whenever they liked and as many times as they liked[55]; however at no other times were they were allowed to move the tiles or gesture as they generated words. Along with the measures of individual differences (described above) the experimental session lasted approximately 35 minutes.

All participants were asked if they consented to be filmed to facilitate later coding and to verify the words produced although the experiment still proceeded if they did not

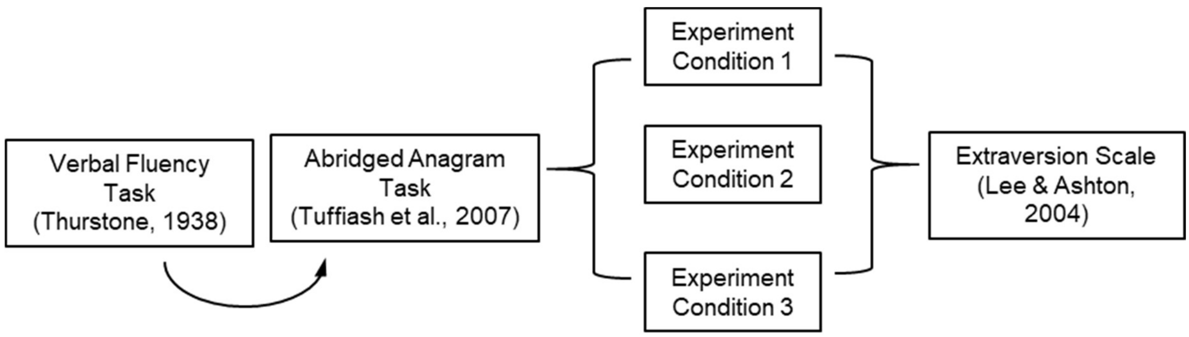
[55] The pilot reported in **Appendix A** explains the rationale for this part of the design.

consent to being filmed. In the high interactivity condition, the participants' level and type of token moving was coded and in the shuffle condition the number and timing of the shuffles was also recorded.

The participants were also profiled along three further measures (a) extraversion and openness to experience (b) verbal fluency and (c) anagram ability. As both the anagram and verbal fluency tests are similar to the scrabble task there was a high chance of it leading to practice effects which would interfere with the counterbalancing of the experimental conditions, so both tasks were used as warm up tasks before the main experimental three conditions. The measure of extraversion was placed at the end of the study producing the order illustrated by Figure 4.1.

Figure 4.1

The Order of the Tasks in Both the Pilot and the Main Experiment. The Numbers Correspond to the Three Experimental Conditions, the Order of Which Was Counterbalanced across Participants



Results

Data availability

The data reported and analysed here are available in **Appendix B** and the long appendices which include second by second detail of the two case study participants summarized at the end of the Results section are available in **Appendix C**.

Aggregate Performance Measures [56]

There were slightly more valid words produced in the high interactivity condition with participants producing an average of 18.35 words in 5 minutes (*SD* = 8.48); in the shuffle condition, participants produced slightly fewer words (*M* = 17.22, *SD* = 6.23) and in the low condition they produced the least (*M* =17.02, *SD* = 6.59). However, in a repeated measures analysis of variance (ANOVA), the effect of condition on performance was not significant, $F(2, 78) = 1.94$, p =. 146 $\eta^2 = .048$. Similarly, there were no significant differences in word length between the conditions: the mean word length in the high condition was 3.68 (*SD* = .212), 3.64 (*SD* = .171) in the low condition and 3.62 (*SD* = .205) in the shuffle condition, $F < 1$. The words produced were marginally rarer in the high interactivity condition as indexed with Zipf scores (high: *M* = 4.172, *SD* = .281; low: *M* = 4.254, *SD* = 2.74, shuffle: *M* = 4.243, *SD* = .234); however, these means did not differ significantly, $F(2, 78) = 1.54$, $p = .220$, $\eta^2 = .038$.

[56] Results prior to exclusions: The means in the three conditions were similar to each other and showed the same patterns as the main data set after exclusions. In the high interactivity condition participants generated an average of 17.83 words (*SD* = 8.34) with an average letter count of number of letters of 3.66 (*SD* = .24) and an average Zipf score of 4.18 (*SD* = .28). In the low condition, they generated 16.88 words (*SD* = 6.52) with an average letter count of 3.64 (*SD* =.18) and an average Zipf score of 4.26 (*SD* = .27). In the shuffle condition, they generated 16.83 (*SD* = 5.94) words with an average length of 3.63 (*SD* = .21) and a average Zipf score of 4.25 (*SD* = .24) A one way ANOVA revealed that the main effect of condition was not significant, *F* (2,82) = 1.38, *p*=. 257, for the numer of words generated, not for the letters, F<1, not for the Zipf score, $F(2, 82) = 1.86$, p = .163.

Individual Differences[57]

Table 4.4 reports descriptive statistics for the measures of individual differences, as well as the matrix of correlations for the measures of performance (number of words produced) and scores on the measures of individual differences (*df* = 38 unless noted otherwise). As expected, there was a strong positive correlation between verbal fluency and anagram skills, $r = .493$, $p = .001$; both were significant predictors of performance in all three conditions (lowest $r = .601$, $p < .001$). Measures of extraversion did not correlate with any measures, nor did openness, with the exception of a negative correlation with word production, $r = .345$, $p = .037$, in the high interactivity condition. The direction of this association is a little difficult to interpret given that openness is sometimes associated with higher self-report measures of creativity, and it may be safer to treat the finding with caution.

Table 4.4

Descriptive Statistics and Correlations Among Measures of Anagram Performance, Verbal Fluency, Openness, Extraversion and Word Production Performance in the Three Experimental Conditions.

	M	*SD*	1	2	3	4	5	6	7
1. Anagram Total	8.90	2.73	-						
2. Verbal Fluency	88.33	25.15	.493 **	-					
3. Openness	50.53	8.56	-.184	-.090	-				
4. Extraversion	52.23	8.49	.133	.074	.345 *	-			
5. High Interactivity	18.35	8.49	.601 **	.717 **	-.331 *	-.022	-		
6. Low Interactivity	17.03	6.59	.679 **	.734 **	-.223	-.047	.819 **	-	
7. Shuffle Condition	17.23	6.23	.630 **	.745 **	-.166	.086	.848 **	.796 **	-

* $p < .05$ level (2-tailed). ** $p < .001$ level (2-tailed).

[57] Individual differences prior to exclusions: Significant correlations were revealed between the anagram score and number of words generated in all three conditions (df = 40; high: $r = .648$, p <. 001,; low: $r = .682$, p <.001; shuffle: r =.645, p<.001) although anagram score did not correlate with the average word length and only with word frequency in the shuffle condition, $r = -.417$, p=.006. Verbal fluency significantly correlated with the number of words produced in all three conditions (high: $r = 0.731$, p <.001, low: r () = .740, p <. 001, shuffle: $r = .761$, p <.001) and with the Zipf score (high: $r = -.477$, p=.001, low: $r = -.341$, p .027. shuffle: $r = -.461$, p =.002). There were no significant correlations with the Hexaco-R measure of extraversion and openness.

Participant Behaviour

An analysis of participant behaviour within the conditions was also conducted. For the low interactivity condition, participant behaviour was controlled so variation was limited but in the high interactivity condition, participants were invited to move the tiles as they wished resulting in a wide range of behaviour. In the shuffle condition, while behaviour was controlled between shuffles, the number and timings of shuffles was under the participants' control.

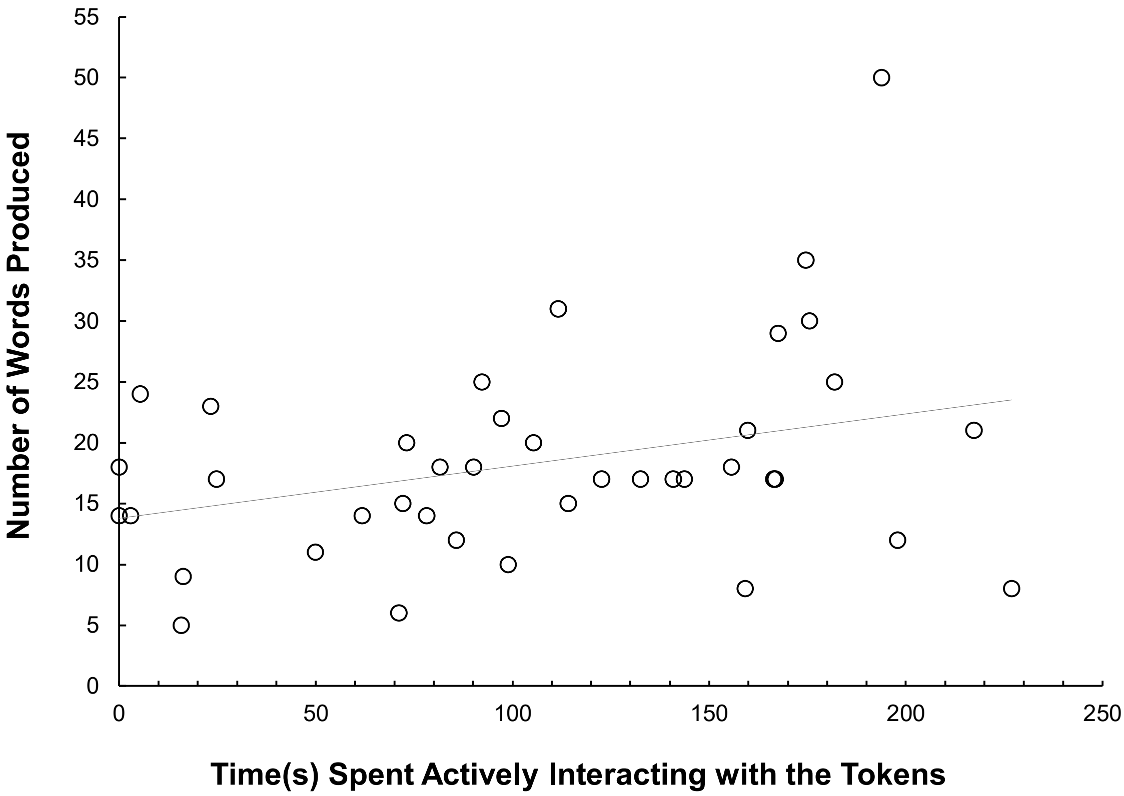
Time Interacting

In the high interactivity condition, the amount of time participants spent moving the tiles was coded using ELAN (https://tla.mpi.nl/tools/tla-tools/elan/). The total time interacting with the tiles was assessed from when a participant touched a tile to when he or she stopped touching it. As there were many moments when a participant touched a tile but did not move it, this was further split into neutral moves (which did nothing to alter the array) and active moves (which changed the array in some way, either deliberate or random). An initial attempt to code the moves as suggested by the pilot into "verifier" or "generator" proved too complex and prone to researcher interpretation. This is discussed further in the limitations section. The split between active and neutral moves provided an objective measure in this situation akin to presentation change and marking in F. Vallée-Tourangeau et al. (2015). Active moves were considered a reflection of interactivity.

The average time spent interacting with the tokens was 106.4 seconds (*SD* = 65.1) out of a possible 300 seconds. Two people chose not to interact at all and from the remaining 38, the shortest amount of time spent moving the tokens was 2.92 seconds and the longest was 226.9 seconds. This suggests that the majority of participants recruit the environment when it is available but the manner in which they do so varies.

Figure 4.2

Number of Words Produced in the High Interactivity Condition as a Function of the Time (in Seconds) Spent Interacting With the Letter Tiles.



There was a significant correlation between the amount of time spent actively moving the tiles and the number of words produced in the high interactivity condition, r = .329, p = .038 (see Figure 4.2); the correlation was marginally more positive when controlling for fluency, $r(37)$ = .356, p = .026. This indicates that the amount of time spent interacting had a continually additive effect contrary to the hypothesis. Also contrary to the hypothesis, the relationship between the amount of time spent interacting and the measures of individual differences was not significant, verbal fluency: r = .117, p = .472; anagram skills: $r = -.021$, $p = .897$. The time spent moving the tiles appears to reflect something beyond individual differences in verbal skills.

To ensure that the total movement time did not reflect an individual difference that would be reflected by an increased performance across all conditions, the correlations between the time spent interacting with the tiles in the high condition with the performance in both the low and shuffle condition were compared. While both were positive, the correlation with words produced in the low condition was not significant, $r = .219$, $p = .175$, nor with words produced in the shuffle condition, $r = .146$, $p = .365$ (and indeed when controlling for fluency these correlations were weaker: low, $r(37) = .199$, $p = .224$ and shuffle, $r(37) = .093$, $p = .574$). This suggests that the time spent moving in the high condition is a unique predictor of the number of words produced in that condition rather than reflecting an underlying individual difference.

Shuffling

In the shuffle condition, the hypothesis that the increased time and cognitive cost would lead to a decrease in the number of shuffles from that reported in Kirsh (2014) was upheld. In the shuffle condition the number and timing of the shuffles was also recorded in ELAN. The largest number of shuffles was 3 with a mean of 1.55 ($SD = 1.13$). Word production performance did not differ as a function of the number of shuffles: 10 participants chose not to shuffle at all ($M = 18.9$, $SD = 6.40$), 8 shuffled once ($M = 15.87$, $SD = 6.12$), 12 shuffled twice ($M = 17.00$, $SD = 6.50$), and 10 shuffled 3 times ($M = 16.90$, $SD = 6.43$); a one-way between-subjects ANOVA with number of shuffles as a grouping factor revealed that the number of shuffles did not have a significant effect on the number of words produced, $F < 1$. Neither verbal fluency, $r = .038$, $p = .841$ nor anagram performance, $r = -.011$, $p = .954$, correlated with the number of shuffles, supporting the results from the high interactivity condition that suggest that changing the array is not related to those individual differences.

Exploratory Analyses

To my knowledge a detailed analysis of behaviour in this task has not been done before. Therefore, a more detailed examination of process was necessary before any further hypotheses could be generated. I first took a subsection of the main sample and subjected these participants to a detailed qualitative analysis. Nine participants were selected for this exploratory analysis. They were selected on the basis of the change in performance from high interactivity to low interactivity: The goal was to use participants who either benefited the most from the ability to move the tiles to generate words, or those who appeared impeded in their ability to generate words in the high interactivity condition. Thus, this sample of 9 included two participants with the highest boost from interactivity: These participants showed an increase in the number of words produced of 18 and 10 respectively. The two participants who experienced the greatest negative impact of interactivity were also selected. Total word production by these participants declined by 7 and 6 respectively. Three participants who showed no change between the conditions were also selected. Additionally, two participants were selected who also showed behaviour different to the overall trend of data; that is that the more time spent interacting with the tiles the greater the word count. One spent over three minutes (181.9s) interacting and yet produced 3 fewer words in the high interactivity condition, the other only spent 5.37 seconds of the whole 5 minute time period interacting with the tiles and yet produced 9 more words in this condition than in the low interactivity. The videos of these nine participants were scrutinized for underlying behaviours that may not be captured by the aggregated means and which indicated underlying strategies in approaching the task. The detailed scrutiny of these nine participants helped us generate three hypotheses which were then tested across the whole sample.

The first aspect noticed was that type of shuffle was not indiscriminate. At times, a shuffle would seed a new idea and word generation would follow quickly, at other time times the participant would seem equally if not more stumped by the array of letters. In other words, shuffles could be lucky or unlucky. A shuffle leads to a random change in the array and this change may quickly generate new words or obscure them. Luck can go both ways. The first hypothesis was that lucky shuffles led to greater word production. As serendipity is the enactment of this environmental luck and is thus relational, this can be evidenced by participant behaviour directly after a shuffle. If the shuffle has been useful, it seems likely that the participant would produce a word directly after. If not, then the shuffle has been less useful in breaking the impasse. Luckiness was therefore indexed as the time taken to produce a word after a shuffle: The faster a word has been produced, the luckier the shuffle.

Second, higher physical engagement leads to a higher overall word production. Engagement here is a more fine-grained concept than time spent interacting with the tiles. Engagement with the environment can be measured by the responsiveness of the participant to the clues thrown up by it: That is, the more participants respond to the environment, the more words they would produce and in contrast, the less they use the environment, the more they will be weighed down by the cognitive cost of movement. This measure was termed the efficiency score–that is, the measure of participants' leverage of environmental opportunities. It was expected that this would be a greater predictor of word production in the high interactivity environment than in the conditions where this strategy is not as easily enacted.

Third, there was an attempt to operationalise 'internal' and 'external' processes. It seems likely that a word verbalised while the participant was not moving the tokens was more likely to come from internal processes, while one which the participant spoke during movements would reflect changes wrought by the array. It was therefore hypothesised that the proportion of words produced while moving the tiles would predict the number of words produced overall in that condition.

In sum, three further exploratory hypotheses were generated:

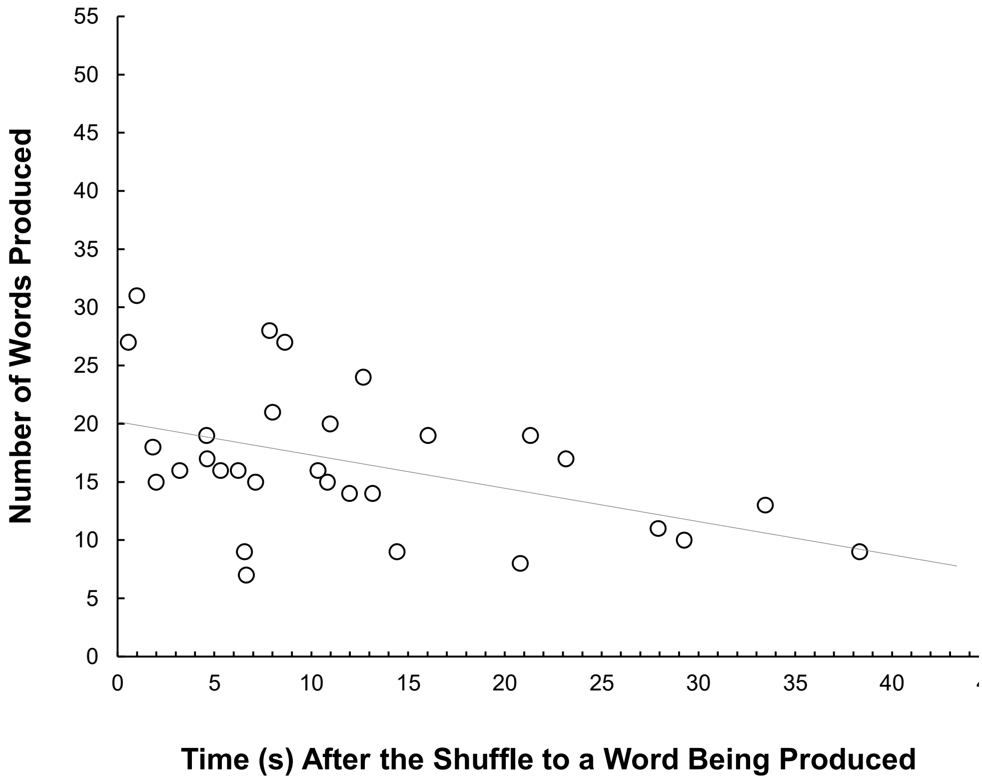
1. Lucky shuffles would lead to a higher number of words being produced overall.
2. Higher engagement with the array would lead to a higher number of words being produced overall.
3. The proportion of words produced while moving the tiles would predict the number of words produced overall.

Luckiness of Shuffles

The time from the end of the shuffle to the production of a word was calculated. The end of the shuffle was chosen because participants' behaviour was once again controlled and their engagement with the tiles limited to what they could manage with restricted movements. Luckiness was indexed as the time from the end of the shuffle to the production of a new word. As expected, there was a wide range in times. Indeed, it was possible for a participant to generate a word while relaying the tiles, the change in the letter array presumably triggering an already liminal word: In this case, a negative latency was recorded, which was the time between the participants uttered the word and the time the final tile from the shuffled set was laid on the work surface. Where participants shuffled more than once the average of the times was taken. An analysis of the correlation between the number of words produced in the shuffle condition and the luckiness of the shuffle indexed in this way was conducted on the 30 participants who opted to shuffle (see Figure 4.3). This revealed a significant relationship between the number of words produced in that conditions and the luckiness of the shuffle, $r(28) = -.460$, $p = .011$. The shuffle represented an element of nonlinear luck that prompted participants' performance beyond their individual skills: The relationship between the luck experienced (as operationalised with the average latency to first word produced) did not correlate significantly with verbal fluency $r(28) = -.316$, $p = .089$, or anagram skill, $r(28) = -.190$, $p = .315$.

Figure 4.3

Number of Words Produced in the Shuffle Condition as a Function of the Luckiness of the Shuffle



Efficiency Score

There are inconsistent benefits to high interactivity. Moving the tiles may not necessarily augment the system's new word affordances and indeed the additional cost may slow the system down if the benefits are not fully realised. In the word production task where the letters are unchanging, a beneficial strategy is to use the same root and change a single letter. In this way a participant may identify the root -ATE in the word set and produce the words D-ATE, F-ATE, G-ATE, H-ATE, L-ATE and so on depending on the other letters available in the set of 7, a switch between state spaces (Maglio et al., 1999). It was hypothesised that an efficient strategy would be easier to follow in a condition where the words are reified physically. It was further hypothesised that it would be a better predictor of performance in the high condition because this strategy could be followed with little working memory cost whereas in a low or shuffle condition the boost from an efficient strategy may be undermined by the cognitive costs required to hold congenial letter arrangements in the head.

I calculated the similarity of the produced word to the word immediately before which I call here the efficiency score. This score assumes that when a participant thinks or sees the word, for example, BREAD, it demonstrates a higher efficiency to remove the B and create READ or remove the A and create BRED than to create an entirely new word. Each word generated by a participant was given an efficiency score. Two scores were calculated: the proportion of letters in the same absolute position[58] and the proportion of

[58] Calculated as the number of letters in the word that were in same absolute position as in the previous word divided by either the number of letters in the word itself or the number of letters in the previous word - whichever was the smallest value - to give a proportion of letters in the same absolute position.

letters in the same relative position[59] . The resulting two proportions give different measures of the similarity of words – it is possible to have words scoring highly in relative position but low in absolute position; for example, if the word READ follows BREAD it scores 0% for absolute position but 100% on the relative position. I, therefore, used the higher of the two proportions as the efficiency score for each word. Finally, the efficiency scores were averaged across participants.[60]

Contrary to the prediction, efficiency scores were similar across all the conditions (high: M= .322, SD = .096, low: M = .316, SD = .089, shuffle: .303, SD = .090), $F < 1$. This indicates that the participants were using broadly the same strategy across the three conditions. I then examined whether there was an effect of using the strategy on word production in each of the conditions. The relationships are illustrated in Figure 4.4. The relationship between the efficiency and the total number of words produced in the high interactivity condition was strongly positive, $r = .597$, $p < .001$; however, the level of efficiency was not significantly correlated with word production in either the low, $r = .223$, $p= .166$ or the shuffle condition, $r = .283$, $p =.076$. This indicates that a good strategy is a significant contributor only in the high condition.

Active Words

It was further hypothesised that in the high interactivity condition those words produced while the participant was moving rather than contemplating the tiles would indicate a higher level of engagement with the envionment (having been triggered by ongoing environmental changes) whereas those words produced after movement would be

---

[59] Calculated as the number of letters in the same relative position (i.e., follows the same letter as in the previous word) and divided by either the number of letters in the word itself less 1 or the number of letters in the previous word less 1 - whichever was the smallest value - to give a proportion of letters in the same relative position.

[60] `I thank Maxwell Vincze for time spent assisting me to develop the efficient score outlined here.

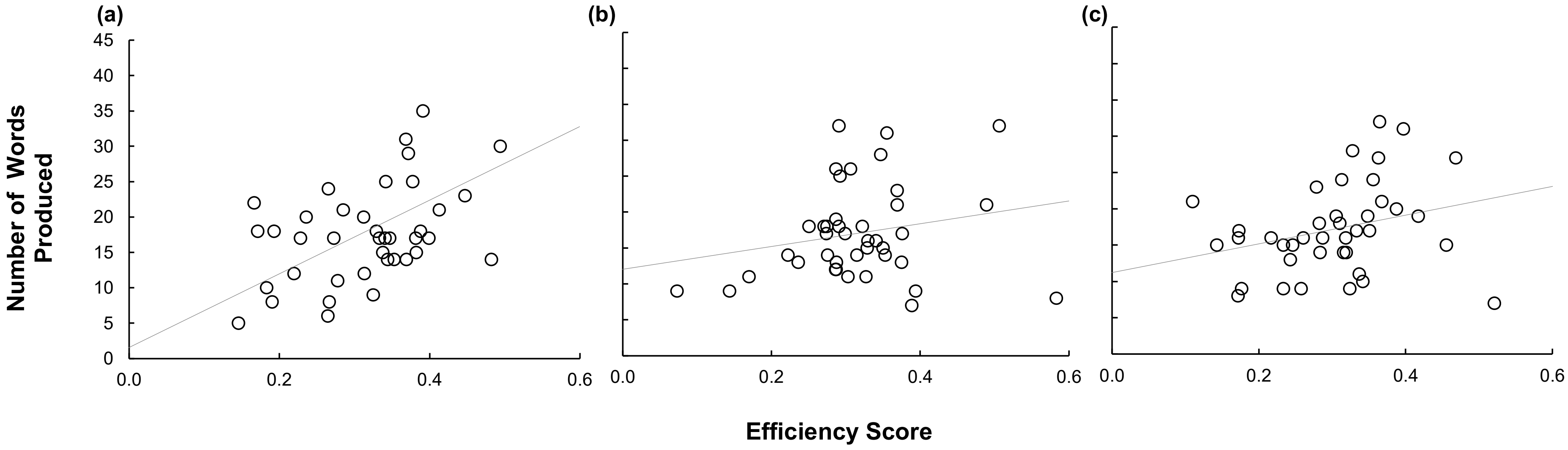
more likely to indicate a word generated from purely internal processes (after contemplation of the array). I therefore calculated the proportion of words announced mid movement and assessed the relationship between this proportion and the number of words produced in the high interactivity condition. A relatively high proportion of words were produced mid movement (M = 43.9%, SD = 25.7%), however, the relationship between this proportion and the number of words produced was not significant, r = .180, p = .266 and the hypothesis was not supported.

Qualitative Analysis

Research in interactivity proceeds from the assumption that including the external world in the cognitive ecosystem augments performance but this implies an optional use of the external world, that a problem solver will recruit the environment when she needs it and rely on her own internal processes to solve the problem when those resources are adequate. Instead, if cognition is regarded as always systemic moments when the external world disadvantages the problem solver also form a part of that system. There is a clear Cartesian dualism to the notion than the “brain” will always “select” appropriate cognitive extensions and scaffolds. Furthermore, problem-solving in a path-rich environment will necessarily result in different routes and strategies, the contingent patterns of which may be masked by aggregate data.

Figure 4.4

Number of Words Produced as a Function of Efficiency Score in the (a) High Interactivity, (b) Low Interactivity and (c) Shuffle Condition .



I therefore selected a participant who did not benefit from high interactivity (Participant 20; P20) and one who did (Participant 41; P41). P41 had the greatest boost from interactivity. Overall, he was higher than average on the measure of verbal fluency: his score was 113 against an average of 88.3 for the sample (+ 0.98 *SD*). P20 was lower than average on this measure, scoring 52 (-1.44 *SD*). Both participants were above average in anagram skill: P41 got all 12 anagrams correct (+ 1.13 *SD*) and P20 got 10 correct (+ 0.40 *SD*).

P41 produced 50 words in the high interactivity condition, 32 in the low interactivity condition and 32 in the shuffle condition. He moved the tiles for an above average time and spent 193.9s interacting with the tiles, compared to the sample average of 106.4s (+ 1.34 *SD*) with over twice as many episodes of activity as average (56 episodes, $M = 24.72$). Indeed, the amount of interaction can be seen in the number of words that were produced during a period of activity. Seventy-three percent of the words generated were produced while moving the tiles. He had a higher-than-average efficiency score in all conditions although it was higher in the high interactivity condition (high = .60, M_{high} = .323, low = .51, M_{low} =.316, shuffle = .40, $M_{shuffle}$ = .303).

P20 produced 12 words in the low interactivity environment and 6 in the high interactivity environment (9 in the shuffle). This indicates that for this problem solver the extended ecosystem was not an aid to thinking, rather it acted as a hinderance. She spent less time than average time interacting with the tiles, 71.23s (- .541 *SD*) However, more of her words were produced during a period of activity (88%). She had a lower-than-average efficiency score in the high (.26) and low (.29) but higher than average in the shuffle (.32)

The following analyses contrast how the singular trajectory unfolded for each participant and specifically assess the coordination of the different systemic elements. The unit of analysis here is not reduced to the individual problem solver but instead contrasts the problem-solving systems formed by the problem solver, his or her environment and the unfolding of the problem over time. In the analysis that follows words produced by the participant are identified with capital letters (e.g., BODE), possible words not produced are written in lower case bold (e.g., **bed)**. Each second of the video is presented in the appendix and the most salient events are highlighted here coded with E and the number to which they refer. The coding referred to here can be found in **Appendix C**.

Inconsistent Effects of Reification

After saying the first word (not a valid spelling), P20 then changed the array to reflect her suggestion. This move was of low utility – the word has already been generated and changing the array did not yield any additional information. Indeed, it is likely to be an impediment to new words because the suggested word is now treated as a unit blocking new ideas. This was a common strategy for P20 who spelt out every word after saying it, that is she would offer a suggestion verbally and then recreate the suggestion using the tiles. This indicates that the direction of cognition was internal to external. Changing the array in this way did not yield any benefit and was only a time and cognitive cost as well as perhaps stifling the generation of new combinations.

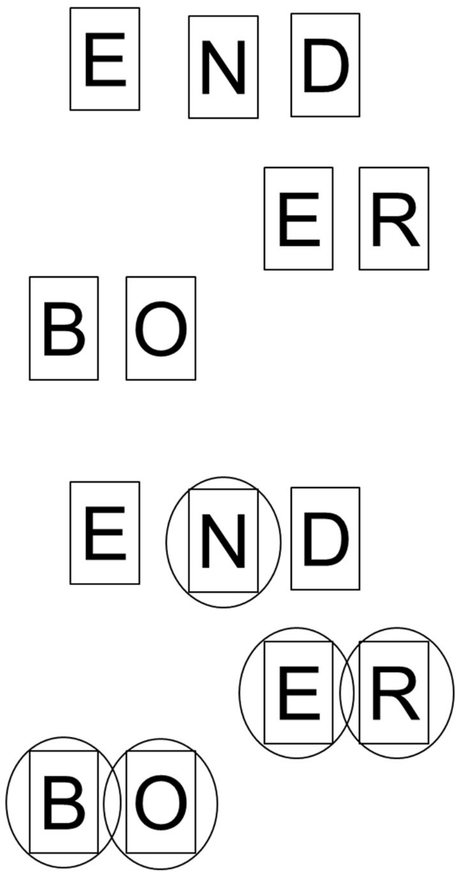
On the other hand, P41 although he reifies his announcements does not leave them to stagnate. He breaks the initial set up quickly, creating a circular arrangement (E5), once the word BOD (E6) is identified this leads quickly to BODE (E9) then BORE (E14), BORED (E17) and BORN (E17) before hazarding a guess at BORNE (E28). This pattern is repeated several times (e.g., ROB (E98) to ROBE (E102) to ROD (E106) to RODE (E110). This pattern of movements would yield a high efficiency score and would be supported by the high interactivity environment because the congenial collections of letters can be reified into a new candidate word. It is noticeable that this is only a useful strategy if the letter tiles are moved sufficiently such as to defeat the anchoring paralysis of a static array. Thus, a level of disengagement is also necessary. Indeed, we can see in E52 the word **bed** is created by the left over tiles but the participant rather incorporates unused letters to keep the array dynamic so this candidate word opportunity is not banked in; rather, he creates BEND (E54)

Microserendipity

The changes to the array yield other left-over collections of letters which are not produced intentionally but are rather an artefact of the movements which have come before. For example, creating the word BODE leaves the letters NRE (E10), later the R and the E become rearranged as artefacts to become ER (E51) which is a much more useful digraph. This movement is not intentional but rather a necessary outcome of a constantly shifting array. For P41, there are several moments when such unplanned moments prompt the following word, which are better understood as moments of microserendipity. The word BRO (E34) leaves the digraph EE on the array. The digraph is incorporated in the following word BREED (E42). In E333 the left-over tiles generated the array illustrated by Figure 4.5: This array leads the participant to suggest the word BONER (E338) then BONED (E342) and finally BONE (E349). This efficient word sequence was sparked by an unplanned change in the array.

Figure 4.5

A Word Discovered by Serendipity. Forming END Led to the Unplanned Array Which Triggered the Word BONER In P41.

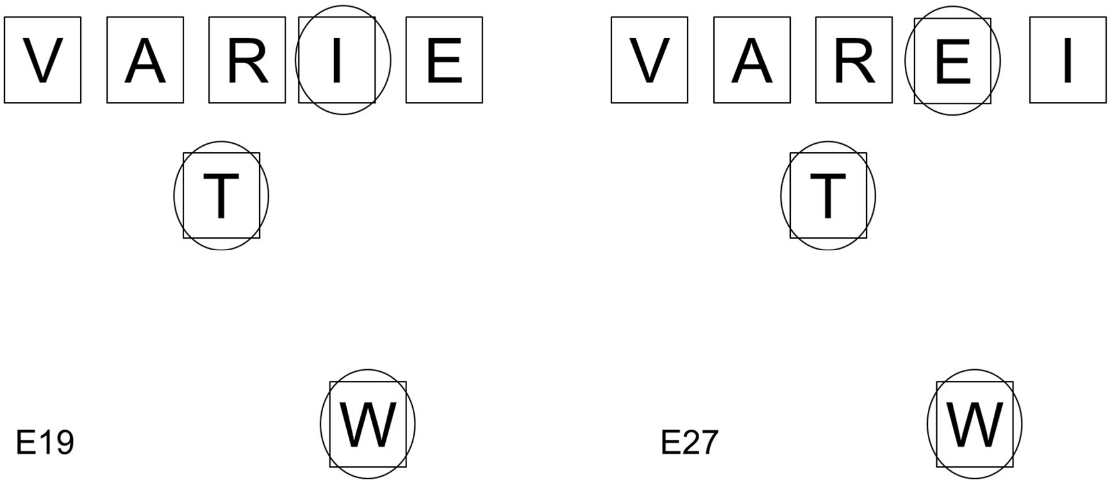


Missed Opportunities

A serendipitous moment can only arise when the environmental luck is capitalized upon by the problem solver. This is something that P20 found particularly difficult. Her lack of movements meant that fewer lucky arrangements were generated but also, she failed to notice others. For example, E19 and E27 offer almost identical letter arrays aside from the rearrangement of I and E (see Figure 4.6) The word WET is spelt out in a triangle in E27 and is prompted, however, this arrangement has spelt out the less common word **wit** previously but this is not noticed by the problem solver (‘wet’ has a Zipf score of 4.67, 'wit' has a Zipf score of 3.67) although being in the same position in E19 as WET is in E27.

Figure 4.6

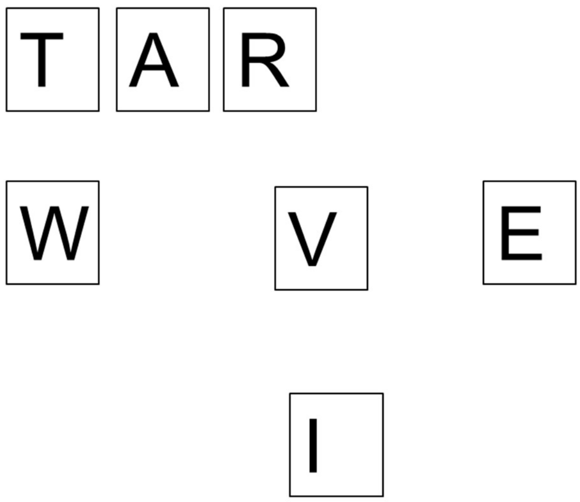
*Contrasting WET and wit Configurations. The First Configuration (E19) Leading to **wit** was Unnoticed by P20, the Second (E27) Leading to WET Noticed and Exploited.*



This underlines the importance of internal and external resources – while the environment may yield a word, if it is not noticed by the problem solver, it will remain inert. This happened often throughout the course of the 5 minutes for P20. It is particularly noticeable when the participant creates **tar** in E185 but does not say the word. This could be for several reasons,perhaps having the word in front of her she forgets to say it, the word does not need to be spoken to take form or because she simply does not notice it (Figure 4.7).

Figure 4.7

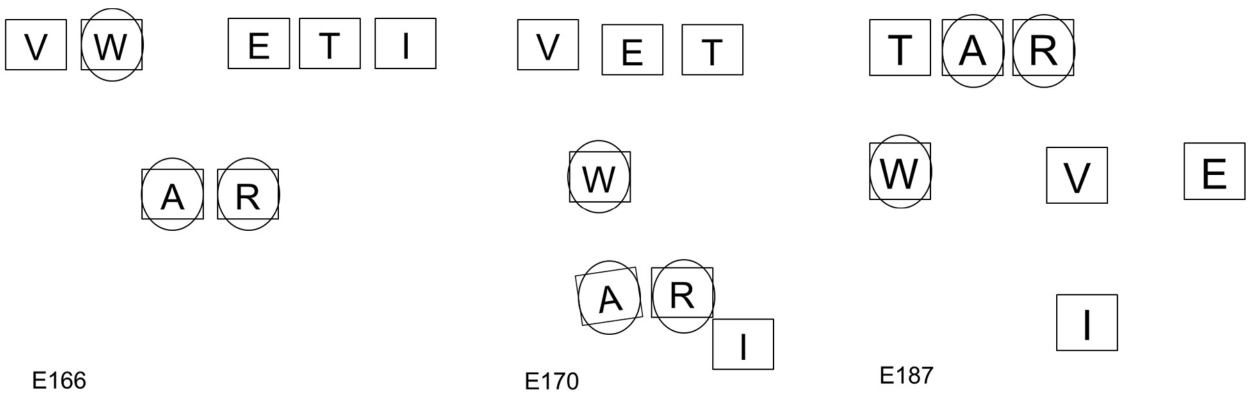
Configuration Spelling ***tar*** *Which Was Unrecognised by P20*



There are other moments of P20 not noticing environmental opportunities. In E71-78 the word **ire** is spelt vertically down and yet not recognised by the participant. The word **war** is made several times by the tiles and not recognised. See Figure 4.8, below.

Figure 4.8

Configurations Leading to ***war*** *Which Were Unrecognized by P20*

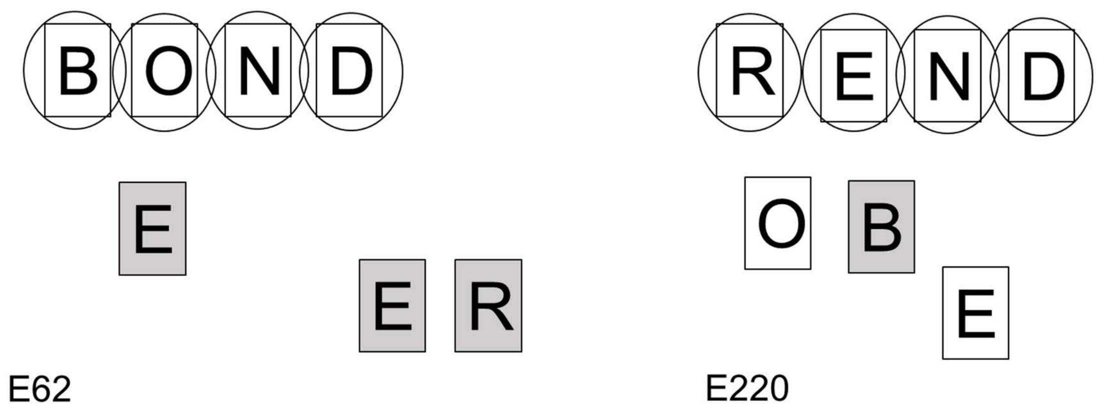


Indeed, any words elicited from the array by this participant were done so in a bottom up or across way aside from this one last word. Those words which were most obvious and yet missed were those which were spelt downward suggesting an inattentional blindness relating to her habitual gaze trajectory.

Although it had been assumed that a high level of engagement would necessarily lead to an improved ability to capitalise on surreptitious candidate words, there are several moments when the array created words which were not noticed by P41 despite his overall high engagement. These were most obvious in the creation of **bond** in E62 and **rend** in E220 the latter of which is not said until E252 as demonstrated in Figure 4.9.

Figure 4.9

*Configurations Leading to **bond** and **rend** Which Were Unrecognised By P41. (The Greyed Out Squares Indicate Those Tiles Which Were Occluded at the Time But Whose Placement Could Be Inferred By Their Previous Position)*



This is in part a function of the kinetics of the task environment, which are hard to capture in the static array presented in **Appendix C**. The tiles were in continual movement and the snapshot reproduced here cannot capture the fluid landscape of letter combinations. It also reflects a level of disengagement (whether a deliberate strategy or not), a reversed sunk cost as it were, quickly disassembling and re-assembling arrays of letter tiles, which, combined with high verbal fluency, prevents output inertia. Finally, it emphasises the contingent nature of serendipity and the inherent difficulty in observing it. In these instances, the environment yields the luck, the participant has demonstrated the ability to recognise those words, but the moment does not happen.

Discussion

Problem-solving emerges from a meshwork of dynamic elements. Resources internal and external to the agent configure a cognitive ecosystem that scaffolds performance. In addition, capitalizing on fortuitous external cues may trigger new ideas and these cues in turn become a constituent part of the cognitive ecosystem. This first study examined these elements to determine how they come into play during a simple word production task. While the broad aggregate scores of the different conditions showed no significant effect of the experimental manipulation on word production performance, close analysis of the video data led me to conclude that participant behaviour did have an effect on the outcomes: The longer a participant interacted with the tiles, the greater the number of words they produced.

Contrary to expectation, the amount of time spent interacting with the tiles was not related to expertise as measured by the fluency and anagram ability. In the literature to date, it has been unclear whether interactivity is used as a prop by those who required more scaffolding in their problem-solving (the idea that cognitive resources will flow to where they are most required) and additionally it was expected that the benefits of interactivity would tail off as people interacted more with the tiles. However, the results suggest a straightforward additive relationship between the time spent interacting with the tiles and the total number of words produced. This suggests that interactivity as a benefit accrues to those who make full use of it and this is regardless of the measured individual differences in verbal fluency and anagram skill. This conclusion was further supported by the significant correlation between efficiency score and performance in the high interactivity condition when such a strategy was easier to enact.

The Benefits of Interactivity

Traditional examinations of interactivity have assumed consistency across participants and across task environments, although the evidence has also indicated that there are significant individual differences in response to an environment that affords interactivity. The change from a digital environment (as in Kirsh, 2014) to the physical problem environment designed in the experiment reported here elicited different behaviours from participants. For example, Participant 41 only chose to shuffle twice in the shuffle condition, but his movements were so quick in the high interactivity condition with arrays assembled and disassembled swiftly that it seems likely that the behaviour here was much like the behaviour of the participants in the shuffle condition in Kirsh (2014) with the added benefit of being able to interact with the chance arrangement.

There are several possible reasons for the increase in performance seen when participants make a fuller use of the option to move the tiles. First, the tiles function as external representations which are simply easier to move and enable participants to leverage their own skills. Morphemically probable units can be formed and manipulated without a resultant working memory load. Second, the movements of the tiles may be unplanned, but the rearranging creates liminal and proto words which are exploited by the problem solver. Third, the movements create left over words which are unanticipated by the problem solver but then recognised by her. The data reported in this chapter show a complex interaction of all three: The internal resources of the problem solver, her embodied behaviour and the environmental affordances that unfold dynamically as the physical model of the problem is modified. A full account of these aspects helps better appreciate their transactional nature.

Interactivity and Microserendipity

Much of the experimental literature examining problem-solving in a situated and movable environment has concentrated on the scaffolding benefits of externalising and reifying representations (e.g., Vallée-Tourangeau et al., 2016). However, this study examined whether chance was also an explanatory mechanism for success in this task.

Chance is inert until it is exploited by an agent. No matter how much luck is present in an environment, it only becomes useful when it is used to produce an outcome (Björneborn, 2017). Serendipity emerges from an interplay of external and internal factors; it is not enough for the environment to create lucky situations, an individual must take advantage of those situations. Equally, an individual's internal resources are not the sole locus of cognitive activity – the opportunities thrown up by the environment are equally important. The central concept of serendipity echoes the central thesis of Kirsh (2014) that chance is beneficial to creativity and problem-solving and suggests that the benefits are not indiscriminate and will accrue to those who are most able to exploit it.

Current research casts serendipity as a phenomenon experienced a posteriori (Martin & Quan-Haase, 2016). This reflects the pure definition of serendipity which suggests that it is the noticing of an unanticipated datum that spawns a new, unsought for solution, explanation or theory (Foster & Ellis, 2014). However, the definition is slippery (Makri & Blandford, 2012) and has been broadened to include both targeted searches which solve a problem via an unexpected route and also untargeted searches which solve immediate problems (Yaqub, 2018). Core to any definition seems to be that it is a problem solved by unanticipated and unplanned means. Using this, the framework of serendipity supports understanding about these unplanned moments are important along the path to a looked-for solution.

Luck, by its nature, is random and when the cost of discarding a shuffle is high, as in the physical environment designed here, the benefits of a lucky shuffle—one which quickly points the problem solver to another word—are outweighed by the high costs of an unlucky shuffle—where the shuffle does not prompt a word in the problem solver. The qualitative, case study analysis demonstrated that luck is often a factor in the high interactivity environment more than a static shuffle environment. As the analysis of P41 demonstrates, this fluid task environment allows the participants the possibility of moving the tiles in a unimpeded manner with a much lower time and cognitive cost than the high interactivity version in Kirsh (2014). This supports the need to fully assess how participants are responding to our experimental manipulations.

A framework of microserendipity allows us to make sense of this interaction between the individual and her environment on a micro level. The detailed analysis of behaviour illustrated by our case studies demonstrated a number of missed opportunities across both those who benefitted from interactivity and those that did not. This underlines the relational nature of serendipity: It is only in the act of noticing (Rubin et al., 2011) that a serendipitous moment occurs. An environment high in lucky affordances does not guarantee that these will be noticed but changing the environment often enough to generate these affordances whether through shuffling on a digital interface or moving in a physical environment increases the likelihood that these affordances will be realised.

Kinenoetic Analysis

The study reported here used a more finely grained method than is usually employed in investigations into cognition—a kinenoetic analysis (KA). This is a necessary development from the theoretical and methodological position mapped out in Chapters One, Two and Three. The analysis here developed from an understanding that the movement of letters in the space did not only represent the experimental manipulation but also an opportunity to track strategy and process. KA requires three things: An instrumentalization of the problem space, behaviours which can be quantitivised and finely grained case studies. The study reported in this chapter illustrates the strength of such an approach. The clear operationalisation of the problem space with letter tiles allowed the problem-solving trajectory to be traced and the relationship between strategy and condition could be clearly evidenced rather than assumed or suggested. The quantitativisation of behaviour allowed a deeper exploration of the underlying parameters across a broad range of participants allowing us to use correlational evidence to spot underlying and emerging trends. Finally, the highly grained case studies revealed the idiosyncrasies of each problem solver's trajectory and highlighted the difficulty of assuming similar behaviours in each experimental condition. When research is interested in the nature of the engagement with the environment, such a method is the only way to fully elucidate the characteristics of that engagement.

Limitations

Measuring Missed Opportunities

The additional complexity of a coupled system in movement makes it hard to create a systematic framework for analysing the number of words missed in a systematic, well operationalized way. A better understanding of microserendipity will likely involve a better understanding of missed opportunities. Missed opportunities were identified as they were seen in the video data but no formal way to categorize a certain un-named configuration of letters as a missed opportunity was provided. Informally, horizontal left-to-right series of letters that formed a word and top-to-bottom vertical series of letters that formed a word, but which was not named counted as a missed opportunity. However, participants demonstrated a considerable degree of flexibility in the selection and combination of letters to produce words: As illustrated in Figure 4.7, P20 produced words in a right to left bottom-up manner, and as illustrated in Figure 4.6, P41 produced a word in a bottom-top-bottom triangular pattern. This open-ended flexibility complicates considerably the formulation of a priori effort to capture missed opportunities. It seems likely that the missed opportunities will reflect both the characteristics of the system and the individual differences of the problem solver (see further discussion in Chapter Eight).

Generating and Verifying

After the initial pilot study, different ways of interacting with the tokens were identified. These were similar to the different behavioral categories picked out by G. Vallée-Tourangeau et al., (2015). It was hoped that a similar type of categorization could be achieved in this study. However, it proved impossible on a practical level to distinguish between different types of behaviour. Over the whole study, in the high interactivity condition, 734 words were produced at a rate of 3.67 words a minute. What might start as a verification would quickly turn into a generation as a novel word stem was identified. Across a data set of 40 participants, it became prohibitive in terms of cost to identify these movements which required a higher level of subjective interpretation. This is part inspired the use of detailed, second by second analysis on a smaller sample, the two case studies.

The initial attempt to code movers as using verifying or generating tactics proved difficult as well because it did not provide an the same objective measure of behaviours as outlined in G. Vallée-Tourangeau et al (2015), rather what was required was the ability to read intention in the material traces. This is impossible and deeply, unhelpfully subjective. This attribute intentionality to material traces requires resisting. It is important to assess the evidence rather than positing unconscious mechanisms (see also Chapter Eight for reflections on analysis of video data).

There was a further attempt to operationalize 'internal' and 'external' processes. One of my exploratory hypotheses suggested that words produced when the tiles are not in motion would reflect internal reflection whereas those produced while the tiles were in movement were hypothesized to reflect 'external' processing. There was no significant effect of internal and external processing in this way but as the coding was taking place it became increasingly apparent that a clear split in this way did not reflect the trajectory of problem solution over time. The word production task is not a binary outcome task and one word reified would spark another, a move in the world may seed an idea in the head. It is not the primacy of either internal or external processes that the behavioral coding and the case studies identified but rather an entanglement of the two. The number of words was an emergent product of internal and external processes and any attempt to separate the two soon became unprincipled.

Shuffle and Interactivity

As discussed in the introduction, the pattern of behaviours evidenced by Kirsh

(2014) is more akin to the flipping of a shape in the game of Tetris. Fast moves of the button allowed the array to be rearranged at will and the main skill then required was the recognition of useful arrays. There was little time or cognitive cost to shuffling. In contrast, the experimental design here loaded the shuffle with time and cognitive cost. Shuffling became a thoughtful action undertaken one or twice (three times at most) over the course of the 5 minutes and so an unlucky shuffle had a greater impact while a lucky one would soon dissipate.

This awkwardness of the shuffle interface meant that the moments of microserendipity were most likely to be seen in the high interactivity condition where moving the tiles quickly could generate unexpected overlaps which were swiftly recognised. For example, one participant created the word DEAF and the left over letter spelled COT, a word which she swiftly stated (Figure 4.10). This was only made possible by an unplanned change in the array wrought by a planned one. Serendipity seems to require an active engagement with the environment and the use of a shuffle condition may not function entirely as Kirsh intended.

Conclusion

The first study of this thesis took a word production task as the base to explore interactions with a material environment and to test the benefits of a more detailed method to understand the how soft-assembled cognitive ecosystems form. The evidence presented here supports the notion that it is not just an interaction with the environment that is important but the nature and quality of that interaction. The work here also presents the first attempt at a more finely grained methodology for capturing the nature of the interaction with the world. The use of mixed methods here – coarse grained quantitative analysis combined with qualitative analysis over the whole state set and even more finely grained case study analysis—fleshed out the insights considerably and offered a clear window onto the nature of the cognitive system configured in terms of dynamic object-thought couplings.

Figure 4.10

Making the Word COT Serendipitously

Chapter Five : Luck in Anagram Solving

This chapter reports the results of a study which built on the findings from the word production task outlined in Chapter Four and which invited participants to solve anagrams. Two pilot studies were conducted to assess anagram solving across three conditions (high, low and shuffle) with both hard and easy anagrams[61]. These pilot experiments (**Appendix D**) generated a series of unanticipated problems which led to extensive reflections on the nature of research in interactivity. The emphasis of the research reported in this thesis then has shifted from establishing the augmentative nature of the artefacts to elucidating the mechanisms through which a system is formed and the manner in which problem-solving is enacted. This translated into changes to the final study. Most important of these was the dropping of the low interactivity condition.

The quantitative results of the final study showed that there was a significant benefit to be being allowed to move the tokens rather than shuffling them. This was in contrast to the initial direction of results in the pilot experiment and either reflects the changes between the tasks or the difficulty of measuring something with a built-in random factor or, indeed, a mixture of the two. Just as in Chapter Four, a finely grained analysis was applied to individual selected cases, this time seven[62]. These analyses suggest an idiosyncratic path to problem solution which is contingent on initial unplanned choices and an active interaction with luck rather than encountering it as a passive recipient. This required me to update the notion of microserendipity developed in Chapter Four. The number of moments of missed serendipity require us to reevaluate a non-agentic account of

[61] The data for these pilot studies can be found in the online appendices
[62] The qualitative analyses for this final study were carried out first on a split sample before being applied to the whole data set

luck that casts the problem solver as passive. There was also discrepancy between observed behavioral markers and both in task and post reports of impasse. The implications of these findings for problem-solving and insight research particularly are then discussed.

Chapter Four supported existing findings that established that interacting with the environment appears to have a positive effect on the number of words produced in classic word production tasks (Fleming & Maglio, 2015; Kirsh, 2014; Maglio et al., 1999; F. Vallée-Tourangeau & Wrightman, 2010; Webb & Vallée-Tourangeau, 2009). The research further supported findings by G. Vallée-Tourangeau et al. (2015) which suggests that it is the quality of the interaction which is important as well as the potential for interaction; in other words, that an experimental manipulation check is necessary to fully appreciate the nature of interactivity. In this way, it followed more closely the second wave of interactivity research which aims to track the quality and nature of the coupling between mind and world (F. Vallée-Tourangeau & Vallée-Tourangeau, 2020). The detailed analysis of behaviour suggested that interactivity may be beneficial because it facilitates more efficient strategies suggesting a recursive relationship between environment and mind rather than a one directional cognitive offload. The study also cast doubt on the ease of observing and sustaining a mind-world boundary where the line between when things are "in the head" and when they are "in the world", instead thought was enacted rather than separated from action. This leads to a recursive and iterative problem-solving process where movement shapes thoughts which enact moves in the world (Bocanegra et al., 2019a)

Finally, the addition of a shuffle condition created a platform to assess the role of luck and randomness. The data suggested that luck was not indiscriminate and that it is possible to experience bad luck, that is that a random shuffle will not have universally positive benefits. The analysis also suggested that the time and cognitive cost of the shuffle meant that moments of "microserendipity" were more likely to be encountered in a high interactivity condition. These moments were described as moments when the accidental and unplanned moves in the environment yielded the answers. The data further demonstrated unpredicted moments of "missed serendipity" that were drawn out in the close video analysis.

Serendipity is a particularly important phenomenon in light of the models of cognition outlined in Chapter One because it requires environmental input and remains underexplored in models which still privilege human agency. As Chapter Three describes, however, it is mentioned anecdotally in many accounts of problem-solving so invites further investigation. With its relational ontology (see Chapter Four), serendipity requires an interaction with the environment because it is necessarily emergent from that interaction and it is, therefore, reliant on process rather than state. In short, it is something that happens in the process of changing the epistemic state of the agent when that state is unstable rather than a description of a stable end state. In other words, it is precisely one of the wider explanatory mechanisms which the thesis intended to explore. Its existence anecdotally and now evidenced empirically suggests that a model of cognition which ignores the environment will be an unsatisfactory one to map cognition in the world; it simply cannot account for this emergent concept. The current study was therefore designed to explore further the role of luck and serendipity in a one solution task, namely anagrams.

Chance, Luck and Serendipity

Chance refers to the continual flow of environmental fluctuations (Griffith, 2010)

Once that chance becomes significant to a person then it becomes either luck or serendipity (Rescher, 1995). Luck in the case of problem-solving describes those things which happen which are beyond the control of the problem solver, but which cause a material change in her environment. There are several different accounts of luck[63] but important for the current study is that all theories maintain a distinction between the problem solver and her environment: "something happens to a person, it is not something that he does" (Griffith, 2010, p. 46). Agency is concentrated in the environment.

Ohlsson (2018) suggests that we need to take into account these moments of external volatility to understand how creative cognition is initiated and argues that creativity (in the form of semantic processing) is triggered by unanticipated, that is a second order[64], change in the environment. It is this volatility that drives the creation of non-routine processes to stabilize the epistemic equilibrium. Uncertainty and unexperienced events require novel cognitive processes. *As soon as agency shifts from the environment to be distributed across environment and person and the result is emergent from accident and action, then the study is no longer one of luck but rather one of serendipity.* The problem solver is not a passive recipient of the answer thrown up by the environment. Material and social worlds are dynamically entwined and produce a rich array of possible perceptions and action affordances (Ross, 2020).

---

[63] There are two main positions which define the class of events to be called luck: the lack of control (Coffman, 2009) and the modal (Pritchard, 2014) accounts. As suggested by the name, the lack of control account stipulates that the event in question occurs when there is no agentic involvement from the person experiencing that luck either in the lead up to the event, or crucially, after. For example, winning the lottery is dependent on the correct numbers being drawn over which the person has no control (Coffman, 2009; Griffith, 2010). The modal account relies on counterfactual thinking: a lucky event is one that could not have happened in the nearest of possible worlds. While the modal account does not expressly require a lack of agency, as Carter and Peterson (2017) argue, the lack of agency is implicit because an agent's actions would be likely to have the similar results in another possible world the only thing which would generate different results is something beyond the agent's control.
[64] This second order is unrelated to second order problem-solving (Vallée-Tourangeau & March, 2019)

In the study reported in Chapter Four, we encountered multiple occasions of "missed serendipity" when the answer seemed to be clearly provided by the environment. These moments can be considered as moments of luck – they are necessarily non agentic and solely situated in the environment. It could be then that it is necessary to enact the luck and actively generate moments of serendipity rather than simply recognise it. For example, Neth and Payne (2011) compared the performance of participants in a yoked coin addition task. The factorial design compared performance in four conditions: when moving and given a random array, when not moving and given a random array, when moving and given the array of a previous successful participant and when not moving and given a successful array. The helpful array is the equivalent of the lucky break, the environment is arranged in such a way as to support and scaffold the problem solver without her doing anything. However, participants made fewer errors when they were allowed to interact with an unhelpful array than when they were given a static already sorted and helpful array, suggesting that the actions themselves may be important to fully leverage the advantages.

The nature of the word production task where the emphasis is on the quantity of words produced means that the effect of luck is somewhat dissipated. A fluid task environment with many more words possible than can be produced in the time means that moments of missed serendipity carry a low cost. As seen from P41, there were still many moments when the environment yielded answers, but these were ignored. An interactive task environment scaffolds a strategy of incremental change – ROSE → RISE →RIDE for example. Shuffling would disrupt this strategy. However, this scaffolding role would be less important in a task with a binary, non-incremental outcome such as an anagram which has only one solution and does not necessarily reward liminal, half formed words. Indeed, these may prove a distraction.

Anagrams

The anagram task has a long history both in recreational problem-solving and as an experimental task. A true anagram is a full word that can be rearranged using all available letters only once to create another word such as EATERS which can be rearranged to spell the word EASTER. More usually though, the letters are presented in a scrambled non word form such as STEAER and the problem solver is invited to uncover the target word from this nonsense letter string.

Performance Moderators in an Anagram Task

The factors that control the difficulty or otherwise of an anagram are complex and multiple and there is little consensus on the features (Knight & Muncer, 2011). Research to date has focused on the difficulty of the target word including its structure, frequency and the age of acquisition of the target word (Gilhooly & Hay, 1977). Alongside the difficulty of the target word, anagrams become more or less difficult when an alternative and incorrect path is suggested by the initial array. Attributes such as the number of moves required, the initial starting letter, the similarity to the target word and pronounceability of the word jumble all have an effect on the ease of solution beyond the characteristics of the target word (Gilhooly & Johnson, 1978; Hunter, 1959). The complexity of the English language means that a systematic system of judging the difficulty or otherwise of anagram arrays is stymied by certain characteristics individual to each word. Most salient of these will be whether the array resembles a distractor word.

In this case, the participant is more likely to become initially stuck on an unhelpful representation and incremental moves will only reinforce the representation. Compare

GRAMANA with ANAMAGR or even AMAGRAN; each arrangement leads to the target word ANAGRAM but the order in which the letters are arrayed influences the difficulty.[65] In the example above, while all the letters are in the right combination the initial presence of the trigram GRA triggers words which start with this particular combination leading the problem-solving to create an unhelpful representation of the problem space which will need to be restructured to support problem-solving. Thus, a change in the array of the same target word can make the same word either easier or harder to solve. The participant in an anagram task who has become stuck will need to restructure the problem more consciously in order to overcome this impasse. To borrow from Steffensen and F. Vallée-Tourangeau (2018, p. 175) there is a "suspended next" which forces the agent to engage effortfully with the problem space. That is, that the next step is no longer clear, and the problem become not solving the problem but rather *how* to solve the problem, what to do next. Thus, it arguably will require more effortful processing – more conscious thought.

Representational Restructuring

There has been a change in the view of the class of problem that anagrams represent. Theoretically, anagrams can be solved by incremental and step by step processing and, in the past, have been often characterized as non-insight problems (Gilhooly & Murphy, 2005). However, it seems increasingly likely that they can also be solved in an insightful way. Certainly, the sort of processing required to solve an anagram in a slow, transformative way is likely to be beyond the capability of most problem solvers within the time usually taken to solve an anagram and this suggests a different route to solution. Indeed, recent normative research indicates that the anagrams elicit reliably high

[65] As we shall see, this statement while seemingly simple becomes fraught with difficulty.

ratings of insight (Webb et al., 2018) and anagrams are now often used as insight problem-solving stimuli (Ellis et al., 2011; Kounios et al., 2009; Valueva et al., 2016).[66]

As outlined in Chapter Two the distinction between insight and analytical problem is not as clearly drawn as might be assumed from a cursory glance at the literature. This is because insight is likely to be idiosyncratic and process driven whereas the problem is itself static. This has led to a move from a focus on insight as a necessary consequence of a hypothesized cognitive task analysis to an affective state, a state which may even reflect complex individual differences as much as the process of problem-solving (Webb et al., 2019). The idea that process and insight as a phenomenological experience are necessarily related is complicated by a research programme that struggles to map the relationship between process and insight in a way which avoids regress – viz. insight is theorized as the outcome of a certain process and so the presence of insight is taken as evidence for the process. While there is no way to track process rather than a post task analysis such a circularity is likely to continue. (The same argument has been made by Ash et al. (2009) as regards theories of representational change.)

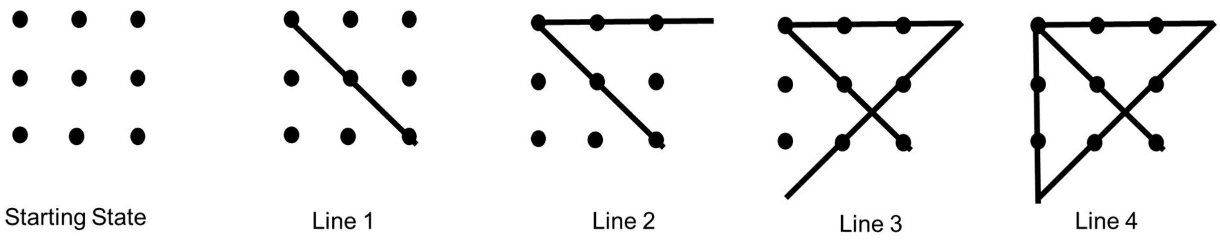
One of the markers of an insight problem which is important to an understanding of how chance may play out in a first order problem-solving task is that it requires *restructuring* (indeed Weisberg [1994, p. 161] is clear that insight refers to "insight brought about through restructuring of the problem situation") after a consciously experienced or not moment of impasse. Take for example, the nine-dot problem. This problem presents the problem solver with a grid of 9 dots and invites them to join all 9

[66] There is an interesting relationship here between method and measuring tool. Neuroimaging research on insight required problem classes (e.g, anagrams, compound remote associates) from which multiple problem tokens are presented as separate trials. This required a focus on repeatability which illuminated the idiosyncratic nature of the classic insight stimuli (see Chapter Six)

with only 4 lines without taking your pen off the paper (see Figure 5.1). The question normally leads to participants staying inside the (imaginary) lines of the square but the solution requires the problem solver to restructure the problem so as to take the lines outside the initial grid (see line 2 where the insight occurs).

Figure 5.1

The Nine Dots Problem and its Solution. The Problem Requires the Problem Solver to Realise That They Can Stray Beyond the Imaginary Square in Line 2.



In two complementary papers, Ohlsson (1984a, 1984b) uses the Gestalt theorists of ideas of restructuring – that is a change in the understanding of the whole problem – and applies them to an information processing model. It is important to note that for Ohlsson this restructuring necessarily presupposes an internal mechanism: “the distinction between the objective situation and observer’s internal presentation of that situation will be presupposed” (Ohlsson, 1984a, p. 71). Ohlsson further cites the high potential physical cost of searching in the flesh and getting it wrong over searching in the head (Ohlsson, 1984b, p. 118)[67]. So, for Ohlsson and in line with the mentalist approach the first principle of restructuring is that "Restructuring is a change in the problem solver's mental representation of the problem." (Ohlsson, 1984b, p. 119). The nature of the processes that underlie the restructuring which is necessary to solve the problem is what drives the main investigations in insight research – are there specific ("special") mechanisms which underlie and generate a sudden change in the problem representation or as these simply the same processes as marks analytical problem-solving (Gilhooly et al., 2015; Weisberg, 2015). This is important because the other difficulties encountered in problem-solving – knowledge deficits and cognitive complexity – have clear causes and solution whereas a faulty initial presentation is more problematic (Ash et al., 2009). Although there is a convergence on the idea that problem-solving of both types is more complex and idiosyncratic (driven in part as discussed in Chapter Three by the qualitative and semi-qualitative investigations into the phenomenology of insight), the field is still dominated by the research which aims to answer that question through psychometric profiling of successful participants (Ash & Wiley, 2006; Chuderski & Jastrzębski, 2018) which suggests that those who are successful on insight problems also have the psychometric profiles for success on analytical problems pointing to a relationship between the two.

Restructuring is conceptually and methodologically problematic as clearly outlined

---

[67] To some extent, this is the stronger argument for an internalist perspective on problem-solving and is underexplored in the literature. It brings to mind Indiana Jones in *The Temple of Doom* presented with the lettered tiles which will give way if he selects the wrong one – incremental action would not be useful here and would likely lead to a collapse. However, the methodological solipsism which suggests that the same isolated and underlying mechanisms of non-enacted thought are at work in that situation as in the situation of a participant faced with a battery of insight tasks for participant credit is slightly farcical. Also, it is important to note the hetereoscalarity of problem-solving – often the physical risks only become apparent after action, prodding the world reveals new knowledge which was not present in the initial cognitive ecosystem and which may go on to inform the decision to take no further action. After all, Indiana discovered the tiles would give way only after making a first mistake. Still, while it seems clear that a strategy of offloading and action may reduce cognitive costs, the proposal that it would not be the best in situations of extreme physical risk is not unreasonable. How many of these we encounter in every day life is perhaps a more salient question.

by Ash et al. (2009). The shift required from one problem space to another requires recognizing that the first problem space is exhausted and an appropriate problem space into which to move. It is still not clear how this state of impasse is reached and recognised (see Chapter Three for a longer discussion this of this). So-called spontaneous restructuring which is not reliant on additional knowledge or external clues is hard to elicit and harder to establish. As I have already reflected, the circularity to establishing whether a phenomenon is the result of a process which requires that phenomenon is problematic. However, there are few methods available to study process if a person is placed in a materially impoverished environment save verbal protocols (Fleck & Weisberg, 2013) or post event interviews (Danek et al., 2014) both of which require the participant to have access to their own processes.

Restructuring and Interactivity

There are two main explanations for the augmentative effect of interactivity. First, the material world can act as a cognitive offload which supports and augments the problem solver's internal resources (Clark & Chalmers, 1998; Risko & Gilbert, 2016; G. Vallée-Tourangeau & Vallée-Tourangeau, 2017) reflecting planned and purposeful moves by the problem solver. Second, these high interactivity environments (whether explicitly conceived as thus or not) allow the environment to play a constitutive role in the cognitive ecosystem through unplanned and unexpected changes (F. Vallée-Tourangeau & Vallée-Tourangeau, 2020). It is the nature of the relationship with unplanned changes that will be investigated in this chapter whereas Chapter Six looks at the relationship with unplanned actions.

Despite Ohlsson's insistence that restructuring is a mental process of search through the problem space until a new problem solution is retrieved[68], restructuring can also occur within the objective and environmentally grounded problem space. This may not be of interest to "special processes" researchers who are looking for the moment of internally driven spontaneous representational change, but it is obviously of interest to those researchers who are interested in the coupling between mind and world. Fleck and Weisberg (2013, p. 452) refer to this environmentally driven cognition as 'data driven restructuring':

> Data-driven restructuring included instances when the individual changed his or her representation of the problem in response to something he or she saw from the physical configuration of the problem […]. Observations occurred as the participant was attempting to construct or implement another heuristic-based solution"

Fleck and Weisberg (2013, p. 253) contrast data driven restructuring with "conceptually driven restructuring" which is driven by a change in the participant's "thinking about the problem goal or constraints" although it is notable that the example they use demonstrates a participant enacting wrong solutions with physical materials so the notion of "concept" as a disembodied thought is somewhat misleading (see F. Vallée-Tourangeau, 2014 for a fuller discussion).

While such restructuring could perhaps be more easily achieved by harnessing the environment (Henok et al., 2018; F. Vallée-Tourangeau et al., 2011), the reification of mental representations could also lead to the crystallisation of unhelpful representations (F. Vallée-Tourangeau et al., 2020). In this way, the role of interactivity in this problem differs quite substantially from the boost to mental workspace (G. Vallée-Tourangeau & Vallée-Tourangeau, 2017) demonstrated by the augmentative role of interactivity in mental

---

[68] The manner of this search is unclear and there is a clear dualism at play here, who or what is generating the search, what defines the boundaries of the search field?

arithmetic (Ross et al., 2020; F. Vallée-Tourangeau, 2013) and the word production tasks discussed in the previous chapter. In both these tasks holding either congenial interim sums or word stems in place enabled greater cognitive flexibility. However, as we have seen anagrams derive part of their difficulty from unhelpful bigrams and trigrams (Knight & Muncer, 2011) which lead to a false representation of the problem. Reifying these will make it more rather than less difficult to restructure the problem helpfully whereas a random shuffle will break up unhelpful bigrams and may lead to a faster solution.

Restructuring and Serendipity

As described in Chapter Three, accident plays an important role in informal observations of problem solvers in first order tasks in the reported literature. Furthermore, it is plausible that the role of accident may be a more important explanatory factor for performance in insight problems than for analytical problems in these experiments. The strong evidence for the benefits of interactivity in analytical problem-solving are not matched by similar evidence in insight problem-solving (see Chuderski et al., 2020; Vallée-Tourangeau, 2014). This is possibly because there is also inconclusive evidence for the role of working memory in insight (Gilhooly & Webb, 2018). If we take an agent centric approach, interactivity is most beneficial as a working memory scaffold and therefore it makes sense that its benefits are somewhat dissipated in insight problem-solving. However, if insight is reliant on taking a path which has not been planned and having a sudden shift in representation, changes in environmental information may as act as trigger for restructuring (Ohlsson, 2018). Thus, solutions may be generated by unplanned changes. On the other hand, to suggest that environmental changes are always unintended is to decouple the problem solver from her environment and the evidence from

Chapter Four suggests that this is not effective. Rather, when given access to a dynamic letter array that array is shifted and moved in tandem with the problem solver rather than at odds with her.

Random shuffling will necessarily be decoupled from internal thought processes and thus, is closer to the pure data driven restructuring described by Fleck and Weisberg where the original plan of the problem solver is derailed by external intrusions. Randomness has been seen to improve performance in a word production task (Kirsh, 2014, although see Chapter Four) where the aleatoric process generates a wider range of candidates to be quickly rejected. Verification carries less cognitive cost than generation, and randomness in this case generates a series of hints which can activate liminal words. Indeed, Kirsh (2014) argues that randomness broadens the search space and is functionally equivalent to an adding a new member to a team. In a similar task, Gavurin (1967) also found a significant boost for a random condition over a static display in anagram solving and further hypothesized that a typical anagram solver may be held back by a strategic approach dogged fixation with pursuing solutions based on her knowledge of English lexical structure without the addition of unplanned and unexpected hints.

However, randomness can be both lucky and unlucky and in the case of anagrams, as with interactivity, a change in the problem array could bring the target word closer or further away. Given that the difficulty of an anagram is in part down to the order of the letters, it is possible, therefore, that altering the array may actually make the anagram harder. It is tempting to think that extending the cognitive system must necessarily strengthen it. This is often the case where the system acts as a simple offload of memory or computation but it is not a necessary quality of the extended mind (Wilson, 2014) and indeed as the research programme into materially and socially extended cognition matures, it is becoming clear that there is not a straightforward additive relationship. This is not to suggest that interactivity is to be disregarded as component of cognition, rather, that a simple linear formula is unhelpful in characterising the cognitive performance emerging from a complex system of internal and external parts. Indeed, that interactivity can sometimes have a detrimental impact on performance but still forms part of the system is perhaps a stronger argument for its constitutive role.

The Current Study

The solution of an insight problem such as an anagram often requires restructuring. This restructuring could either be triggered by internal or external changes in the problem space. The current experiment aims to assess to what extent such restructuring could be facilitated by interacting with the world in two ways: first with the human as primary agent and second, by casting the environment as primary agent by introducing a random condition. The introduction of a condition where the change of the array is not under the control of the problem solver allows us to begin to disentangle the coordination of person centered and environmental agency. Both conditions can be seen as different levels of interactivity: controlled (or person initiated) and uncontrolled (or environment initiated).

Of course, this is an artificial, porous and messy boundary: the high interactivity condition will necessarily encompass both person and environmental agency. An initial problem-solving attempt will change the array in line with the conceptual thought processes but this change is likely to then drive further changes in a cyclical manner prompted by the change in the physical presentation. Equally, the shuffle condition is unlikely to solve the anagram entirely (although this is possible, 1 out of the 125 possible combinations will be the correct answer) and a changed array will only be successful if it prompts reflection on the part of the problem solver. A system is necessarily complex and an examination of the way the parts move will only be valid if it assesses them in motion.

Pilot Task

A technical pilot was first run as two separate experiments. Both experiments used high and low interactivity conditions alongside a shuffle condition as in Chapter Four but the nature of the shuffle varied between them: Free shuffle where participants could shuffle at will and prompted shuffle where participants were invited to shuffle every 20 seconds. Thirty participants took part in Experiment A (Free Shuffle) and 29 participants took part in Experiment B (Prompted Shuffle). The data and a description of the experimental methodology of the pilot tasks can be found in **Appendix D.**

The pilot task was non-conclusive. Experiment A indicated that while people tended to do better in the shuffle condition and worse in the low interactivity condition, this was not significant. Experiment B also showed a boost for the shuffle condition and this was significant when high and shuffle conditions were compared. This supports the initial hypothesis that a shuffle would provide a useful hint and introduce novelty into the cognitive ecosystem. However, in line with the overall research question which intends to assess what people actually do, further quasi experimental conditions were created and participants were assigned to these artificially created conditions on the basis of their behaviour. This yielded seven conditions: (a) a genuine high condition (when the participant moved the tiles in any respect[69]), (b) a non moving high condition (when the participant was assigned to the high interactivity condition but did not move the tiles), (c) a

---

[69] It is not intuitive to cluster moving just one tile with moving many tiles but to create these conditions, a cut off point (however unsatisfactory) was required. In the main study reported the time spent moving the tiles was also assessed to offer a more granular measure.

low interactivity condition, (d) a true free shuffle (when the participants opted to shuffle), (e) a non-shuffling free shuffle condition when the participants were assigned to the free shuffle condition but opted not to shuffle, (f) a true prompted shuffle condition and (g) a non-shuffling prompted shuffle condition.

When actual behaviour was assessed this effect described above was reversed and those who actually moved the tokens in the high condition did better than those who actually shuffled in the forced shuffle condition albeit an advantage that was non-significant. In general, taking behaviour into account, participants were better off not shuffling, reaching near ceiling when they did not move in the high, prompted shuffle and free shuffle conditions.

However, this result does not take into account the reasons for participants behaving in certain ways: Observations during the experiment and while coding the video suggested that participants would initially attempt to solve the problem without moving the tokens. If they solved it in this initial period they would not *need* to move the tiles, if however, they failed to solve it they would opt to move the tiles or have reached the point where they are prompted. Therefore, the direction of the effect could be reversed. It is not that not moving led to success, but that success led to not moving. It is for this reason that assigning people to conditions based on behaviour would be problematic.

Nevertheless, perhaps it is possible to just exclude those who did not move the tokens from the shuffle and the high conditions. From this perspective, that is taking into account all the "true" experimental groups the low interactivity condition appeared to be the one in which participants performed the best. However, if lack of movement indicates finding it easier then excluding those people from the high and the two shuffle conditions but not from the low condition means this latter condition would be artificially inflated by those who found it easy but were retained—it is impossible to exclude those people who did not need to move the tiles in the low condition because there is no behavioral evidence for that.

To date there have been few manipulation checks on experimental work in interactivity (the experiment in Chapter Four among the first to do so). It is assumed that people adhere to the experimental instructions in a uniform way despite it being this very behaviour which is of interest and supposedly driving the difference in performance across conditions. It is only the low interactivity condition which is controlled: hands are normally held flat on the table. The high interactivity condition can involve a range of movements over objects from no moves at all to a continual engagement with the movable objects[70]. Yet when this is taken into account such as in Maglio et al. (1999) or the previous chapter it becomes clear that the level of engagement with this manipulation is key. There is an echo of the assumptions that are being unpicked in insight research. Just as it is increasingly being realized that insight is not guaranteed simply from the problem, interactivity cannot simply be assumed from the condition. On the other hand, people cannot be divided by behaviour because of the complexity outlined above. The second wave of interactivity research requires engagement with these issues in order to progress.

Once we move beyond the dualistic notion that the environment only acts to augment normative cognitive processes to a vision of cognition as being inherently enmeshed then the idea of comparing a high and low condition become less important. The focus becomes not whether interactivity augments cognitive processes but rather how interacting with the environment shapes forms so called transient extended cognitive

[70] And indeed other actions as we shall see in Chapter Six

systems (Wilson & Clark, 2009; Chapter Two) and how these shape the underlying problem-solving processes. A low interactivity condition does not yield any information about this research question. Furthermore, evidence for strategies can only be inferred rather than directly traced as they can be from the movement of the objects (as illustrated in the case studies in Chapter Four). Therefore, as the study reported in this chapter did not aim to assess whether a high interactivity environment augments anagram solving but rather to unpick potential mechanisms, a decision was made to drop the low interactivity condition. This yields two main benefits: First, those who do not engage with the experimental manipulation can be excluded without disrupting the balance between the conditions, second, without being primed with a condition which expressly and artificially restricts movement, it is clearer to assess whether people do naturally engage with the environment. This is a claim often made in the literature but rarely tested.

Experience and Impasse and Interactivity

When addressing the question of why someone might choose to move the tokens in either the high or the shuffle condition, it seems plausible that this might be more likely to occur when they encounter an impasse, that is they get stuck on the problem (See Chapter 3). The pilot task featured three possible moving conditions across the two experiments: high interactivity, shuffle (across half the participants) and prompted shuffle at 20s and 40s (across the other half). The means for the proportion of correct answers and the latency to solution below would support the hypothesis that interactivity occurs (whether by way of a shuffle or moving) when encountering impasse: Those who did not move at all in the conditions which allowed movement, solve the anagram faster and performed at ceiling for correct anagrams.

Informal observations suggested that the participants' behaviour when they reached what the researcher presumed to be impasse was important, but this was not formally measured in the pilot. Impasse involves the feeling of being stuck. Ohlsson (1992, p. 4) describes it as "a mental state in which problem-solving has come to halt". It seems plausible that the anagrams that were solved over a longer latency involved a moment of impasse. It is also plausible that moving the tokens whether by shuffling or by moving them around in order to restructure the perpetual problem space and unveil more opportunities, indicate the participant has exhausted the current problem space, in other words has encountered impasse. Conversely, solutions arrived at before moving or shuffling are ones where the participant has not reached impasse. Measuring impasse in the main study will allow this hypothesis to be tested.

The Main Study

The current study was designed to test whether a model based on luck or one based on serendipity captures better the way problems are solved. The relatively high numbers of participants in the pilot studies reported above allowed some level of confidence in the predictions and so current study was pre-registered prior to data collection. The preregistration can be accessed here: osf.io/7pkzt.

Much like Gavurin (1967), I hypothesised that the benefits of random suggestion and the usefulness of new bigrams would lead to more anagrams being solved (and indeed selected an anagram as a promising platform to increase the likelihood of luck playing a key role). I further hypothesized that randomness would be particularly useful when the participants had encountered impasse and were 'out of ideas' by generating novel ideas in line with Kirsh (2014). In this way the shuffles would act as hints which have been shown to be beneficial in insight problems (Ball & Litchfield, 2013; Chuderski et al., 2020).

In order to repeat the analysis in Chapter Four, I was also interested in the extent to which the participants chose to interact with the tiles. To solve the problems outlined above, I had only two conditions (a high interactivity and a shuffle condition) and omitted the low interactivity condition—where participants are prevented entirely from moving the tiles–which acts as a typical control condition in experiments in interactivity. I preregistered the creation of a third post hoc condition of non-movers if enough participants opted to not move the tiles. This allowed me to remove those who failed the experimental manipulation (i.e., did not shuffle or move the tokens) without unbalancing the conditions (as discussed above).

Method

Participants

Confirmatory Sample Participant Characteristics

The final confirmatory quantitative sample consisted of 64 participants (F = 55) with ages ranging from 19 to 55 (M = 25.19, SD = 8.68) who were recruited in return for course credits.

Exclusions

Eighty participants, representing a mixture of undergraduate and post graduate psychology students, were recruited in return for course credits. All participants had performance data collected and were analysed at a coarse-grained behavioural level (e.g., time spent moving the letter tiles). The more finely grained qualitative analysis was used on a nested sample using an opportunistic case analysis strategy informed by analytical notes taken during the preliminary behavioural coding. Seven videos (10% of the data set) were randomly selected after the quantitative data analysis for the qualitative phase.

Excluded Participants. 10% (8) of the initial participant pool were selected at random[71] and constituted the exploratory sample (Anderson & Magruder, 2017). These were then excluded from the final analysis leaving 72 participants in the confirmatory sample. Eight further participants were excluded from the confirmatory sample: five for video data loss, one for consistent failure to follow instructions, one who became visibly became upset during the verbal fluency task (explained below) and one who did not adhere to the instructions requiring English as a first language.

Excluded Trials. Six trials were excluded either because of experimenter error or because the participant was not following instructions. Commensurate with the preregistration, a further 20 trials which involved participants solving the anagram in under 3 seconds were excluded. Trials were also excluded from the final data set if the participants did not move the tiles. This is more fully outlined in the first section of the results. None of the trial exclusions resulted in a participant being fully excluded.

Design

The participants were invited to solve eight anagrams: four in a shuffle condition where movement was restricted to picking up the tiles, shaking them and then randomly relaying them in a straight line and four in a high interactivity condition where they were able to interact with the tiles in any way they chose. The interactivity level (high, shuffle) was a repeated-measures factor. The anagrams were presented in four blocks of two

---

[71] using the RANDBETWEEN function in Excel

anagrams alternating between the conditions and the order of the initial condition was counterbalanced across participants.

Data Types

There are three main data types analysed in this study. First, the aggregated mean performance (proportion of correct anagrams and latency to solution) across the participants and measures of self-report (feelings of insight) data. Second, behavioural data which were produced by observation and then quantitativised using the methods of interaction analysis (Bakeman & Gottman, 2009; Bakeman & Quera, 2011). In this instance, the coding schemes were pre-specified after watching the split sample dataset. Finally, evidence from video based finely grained analysis is used to elaborate on the causal mechanisms for the results evidenced in phases 1 and 2 of the data analysis. These data have therefore been produced using a quantitative-qualitative approach (Onwuegbuzie & Combs, 2011).

Materials

The anagrams were selected from the set of hard anagrams presented in Novick and Sherman (2008). To ensure that the target words were well known among the participant population, all 20 target words were presented to a sample of 25 undergraduate students from the same participant pool (who did not go on to take part in the full study). The participants were asked to rate their familiarity with the words on a Likert-type scale of 1 - 5 where 1 was "not familiar at all" and 5 was "very familiar". These scores were then averaged to create a familiarity index. This led to the selection of eight most familiar target words which were presented in the same initial array to each participant, see Table 5.1. All were considered to be familiar to the target population. The order of the target words was counterbalanced across conditions.

Table 5.1

The Target Word and the Presented Array.

Target Word	Familiarity Index	Presented Array
Power	4.96	WROPE
Madly	4.84	MYLAD
Smoky	4.80	MYKOS
Minus	4.72	SINUM
Milky	4.72	LIMYK
Naive	4.64	VANIE
Gnome	4.32	MENGO
Probe	4.04	BROEP

The participants were also profiled along two dimensions of individual differences, anagram expertise and verbal fluency. Their level of anagram expertise was assessed using the questions in Novick and Sherman (2003) which required them to state how often they solve anagrams (every day, a few times a week, once a week, a few times a month, once a month, a few times a year, never) and how they would rate their anagram expertise in relation to other students (on a scale of -5 to +5 with 0 being average). Following, Novick and Sherman, the frequency was converted to a 1 -7 scale and two items collapsed to create a composite experience (Cronbach's $\alpha = .752$). The participants were also profiled on their phonemic verbal fluency using the F-A-S test and their semantic verbal fluency using the ANIMALS test (Whiteside et al., 2016). Both required participants to say aloud, over the course of a minute, words that fit the given category. The ANIMALS test required them to name animals, the F-A-S test words which begin with the letters F, A and S. The total number of words from all four trials were added together to yield a global verbal fluency score.

Additionally, if a participant solved an anagram successfully, they were asked to answer a series of insight questions (taken from Webb et al., 2018). This required them to rate using a Likert-type scale ranging from 1-100 where 1 = low and 100 = high : (1) the confidence that the given response was correct ("very unsure" to "very sure"), (2) the strength of the aha experience ("very weak" to "very strong"), (3) the pleasantness of the insight experience ("very unpleasant" to "very pleasant"), (4) the surprising nature of the insight experience ("not surprising at all" to "very surprising"), and (5) the feeling of impasse before the insight experience ("no impasse at all" to "very stuck").

Procedure

Data were collected in a purpose build observation laboratory with three cameras and answers were recorded on a Qualtrics interface (www.qualtrics.com). First, participants were invited to complete demographic information which included the measures of anagram expertise. Before being invited to solve the anagrams, participants were asked to solve two practice anagrams: one high and one shuffle. If they did not opt to shuffle in the practice condition, they were invited to practice shuffling as a stand-alone procedure.

The anagrams were presented in the form of movable tiles (2.7cm x 2.4cm) taken from the word game Appletters™ . They were presented on a small tray behind a screen and latency was recorded from the time the screen was removed (see Figure 5.2). The eight anagrams were presented in four blocks of two anagrams in the high interactivity and in the shuffle condition. The anagrams were counterbalanced across the conditions and the order of the conditions was counterbalanced across participants. Participants were given 90 seconds to solve each anagram

Figure 5.2

The Experimental Set-up. Particpants Were Presented the Anagrams on a Tray to Maintain Consistency Across Trials With the Movable Artefacts.

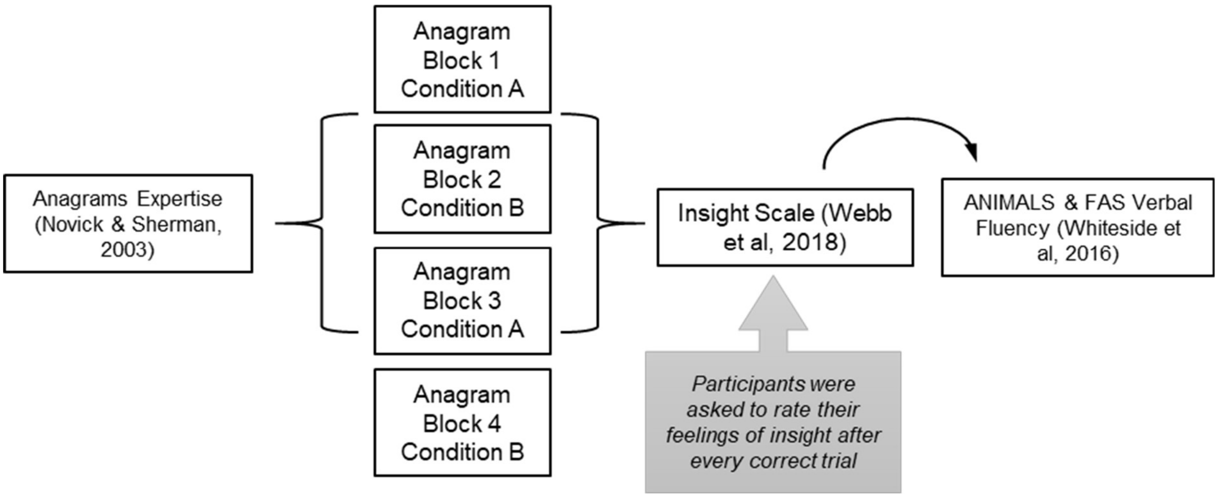
. If the participants did not solve the anagram in 90 seconds, they were told the answer. Before each trial they were reminded of the experimental condition and the requirement to announce impasse using the following wording (adapted from Fedor et al., 2015):

> You may know the answer straight away but you may also get stuck, we would therefore like you to indicate when you are stuck, that is when you have exhausted all the solutions that initially come to mind. Please do this by saying, 'I'm stuck'. You can do this as many times as you like. You may then carry on playing with the tiles to solve the problem.

If they solved the anagram, they were invited to rate their feelings of insight. Finally, participants were invited to do the verbal fluency tests: The F-A-S was administered first followed by the ANIMALS. The procedure is illustrated in Figure 5.3.

Figure 5.3

The Experimental Procedure. The Anagrams Were Presented in Counterbalanced Blocks of Four Anagrams and the Order of the Conditions Was Further Counterbalanced



Quantitative Results[72]

Data availability

The data reported and analysed here are available in **Appendix F** and examples of the coding Participant 10 is available in **Appendix G**.

Non-moving Trials

In order to elicit natural behaviour, participants were told that they were under no obligation to move the letter tiles. However, a trial where the tiles were not moved does not reflect the experimental manipulation (as discussed above) so therefore, an additional 128 trials were excluded because the participants did not move the tiles. No single participant

[72] Post hoc power analyses were conducted because the final sample size could not be determined before data analysis but power analyses conducted a priori are outlined in the pre-registration.

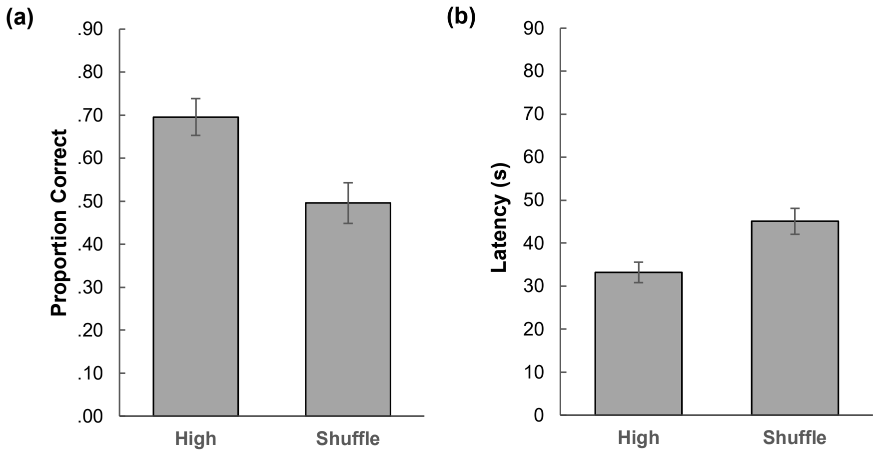
was excluded because each participant moved the tiles in at least one trial. However, six participants opted to not move the tokens at all in one of the two conditions. and so were excluded from the between conditions comparison although they were retained for within condition correlational analysis. Eighty-six of these trials were from shuffle conditions and 42 were from the high condition. Of these trials, the overwhelming majority were correctly solved: all of the trials in the high interactivity condition were solved and only 4 of the 86 (4.7%) shuffle trials were not solved. These trials were also solved quickly: In the high interactivity condition, the average latency to a correct solution was 18.01s (SD = 16.09) while in the shuffle condition it was 19.16s (SD = 15.70). That no one participant opted to not move the letter tiles at all was a clear support for our initial hypothesis that participants would opt to move when given the opportunity. The high success rate and low latency to solution would suggest that these were trials where the participants did not require external support to scaffold their answer as suggested by the informal observations from the pilot.

Performance

Fifty- eight participants had trials remaining in both conditions and so were retained for the within-subjects comparison of performance in the two experimental conditions. Contrary to the hypothesis, participants solved more anagrams when they were allowed to interact with the tiles freely (see the left panel of Figure 5.4). Participants in the high interactivity condition achieved an average proportion correct of .70 (SD = 0.32) whereas in the shuffle condition only half the anagrams were correctly solved, .50 (SD = 0.36). This difference was statistically significant, $t(57)=4.13$, $p < .001$, Cohens's $d = .587$ According to a post hoc power analysis the current sample size yielded 99% power to detect this effect size.

Figure 5.4

Anagram Performance. (a) Proportion of Correct Trials, (b) Latency to Solution (in Seconds) in the High Interactivity and Shuffle Condition. (Error Bars Represent Standard Error of the Mean).



Forty-three participants solved at least one anagram in both the conditions which allowed the comparison of the latency to solution for correctly solved anagrams. People solved the anagrams faster in the high interactivity condition: The latency to solution was lower in the high interactivity condition (*M* = 33.18s, *SD* = 15.67) than in the shuffle condition (*M*=45.09s, *SD*=19.78) as illustrated by the right panel of Figure 5; this difference was statistically significant, $t(42)=2.87, p=.006$, Cohen's $d = .670$. A post hoc power analysis suggested 99% power to detect an effect of this size.

Post Impasse Performance

I was particularly interested in what occurs when a participant encounters impasse. Despite the results of the post-trial questionnaires on insight which indicated a difference in feeling "stuck", participants reported impasse during the trials at broadly the same rates. Of the 58 participants with trials in both conditions, in the high interactivity condition the mean proportion of trials in which the participants reported impasse was .38 (*SD* = .33) and in the shuffle condition it was slightly higher, .46 (*SD* =.37); this difference was not statistically significant, $t(57) = 1.53$, $p = .132$, Cohen's $d = .23$. I also looked at the average time to declaring impasse across the two conditions: participants were slightly quicker to declare impasse in the high interactivity condition (*M* = 24.87s, *SD* = 16.41s) than in the shuffle condition (*M* = 26.09, *SD* =17.65) but again this difference was not statistically significant, $t < 1$. Thus, it seems likely that any difference in performance in these conditions where participants declared impasse was not driven by participants being quicker or more likely to declare impasse in any one condition.

Thirty-five participants declared impasse in both conditions allowing us to compare performance after impasse. Again, contrary to our hypothesis, those participants who declared impasse went on to perform better in the high interactivity condition with an average proportion of correct trials post impasse of .47 (*SD* = .45) whereas the average proportion of correct trials in the shuffle condition after impasse was .19 (*SD* = .33). The differences between these two conditions was significant, $t(34) = 3.24$, $p=.003$, Cohen's $d = .71$. A post hoc power analysis suggested the test had 98% power to detect an effect of this magnitude. Only 10 participants solved the anagrams in both conditions after declaring impasse. They were slightly faster in the high condition ($M_{High} = 51.64$s, *SD* = 16.00) than in the shuffle condition ($M_{Shuffle} = 52.34$, *SD* = 25.35) but this difference was not significant, $t < 1$.

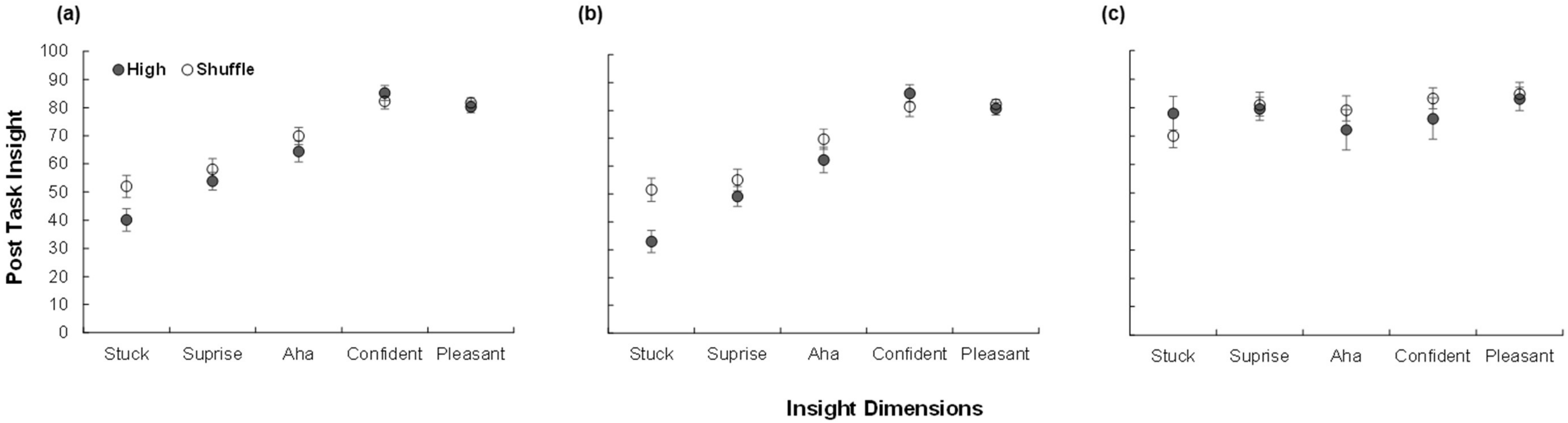
Insight

The levels of reported insight were broadly similar across both conditions as illustrated in Figure 5.5a. Again the data of the 43 participants who correctly solved an anagram in each condition (and therefore reported insight) were analysed. They were more likely to report higher feelings of being stuck in the shuffle condition (M_{High} = 40.12, *SD* = 26.12, $M_{Shuffle}$ = 52.02, *SD* =25.99), $t(42) = 2.68$, p=.011, Cohen's *d* = .46 .The differences across the other dimensions were not statistically significant (smallest non-significant *p* = .146 for the aha dimension).

As a pure insight sequence is theorised to require impasse (Ohlsson, 1992), I then compared the ratings of insight for those who declared impasse and those who did not (impasse and non-impasse trials respectively). Thirty-three participants, the overwhelming majority, had non-impasse trials allowing us to compare their answers (Figure 5.5b). There was a significant difference again for feeling of impasse participants reporting higher rating of feeling stuck in the shuffle condition (*M* = 51.51, *SD* = 23.66) than the high condition (*M* = 32. 98, *SD* =22.63), $t(32) = 4.20$, $p < .001$, Cohen's *d* = .97. All other differences were not significant, (smallest non-significant *p* = .111 for the aha dimension). Only 10 participants had impasse trials across both conditions (Figure 5.5c) and there was no significant difference in any of the dimensions (smallest non-significant *p* =.242 again for the aha dimension).

Figure 5.5

Insight Ratings Across the Five Dimensions for both the High and Shuffle Scores (a) Over the Whole Date Set (b) When Impasse Was Not Declared (c) When Impasse Was Declared.(Error Bars Represent Standard Error of the Mean)



Behavioural Analyses

I next looked at how participants behaved in each of the conditions. Participants engaged faster with the tiles in the high condition. The average latency to first move was 10.02s (*SD* = 8.56) in the high condition whereas in the shuffle condition participants shuffled first after an average of 17.06s (*SD* = 13.94): this difference was significant, $t(57) = 4.11$, $p < .001$, Cohen's $d = .60$. However, while this might indicate that interacting with the tiles was an important driver of solutions the proportion of time spent actually moving the tiles in the high interactivity condition showed no relationship with the proportion of correct answers, $r(61) = -.06$, $p = .645$.

Individual Differences

The participants were also profiled on their verbal fluency and a self-assessment of relative anagram expertise. It is worth noting that the experiment was not powered to adequately test these correlations and they are therefore presented here as exploratory analyses. Verbal fluency was calculated as the total of the three phonemic and one semantic fluency task. The average number of words produced across these conditions was 60.81 (*SD* = 16.81). Counterintuitively, verbal fluency did not significantly relate to performance as measured in terms of the proportion of anagrams solved in either of the two experimental conditions. In the high condition there was no relationship at all, $r(62) = .07$, $p = .581$, while there was a slight positive correlation with performance in the shuffle condition but this was not significant, $r(57) = .158$, $p =.233$. There was also no relationship between verbal fluency and latency to solution in the high, $r(55) = -.06$, $p = .661$, but there was a significant relationship between speed to solution and verbal fluency in the shuffle condition, $r(53) = -.283$. $p = .036$. Verbal fluency did not correlate significantly with the time spent moving, *r(*62) = .130, *p* = .311 in the high interactivity condition, nor with the rates of impasse in either condition: in the high interactivity condition, there was a slight positive correlation with verbal fluency, *r*(61) = .115, *p* = .370, while this was negative for the shuffle condition *r*(57) = -.105, *p* = .427[73].

The relative experience of the participants was also assessed. Eight participants did not answer this question so the expertise of 55 participants was recorded. The average level of the composite score was 0.37 (*SD* = 2.72) indicating that participants rated their expertise as below average. Indeed, this level corresponds to the less skilled solvers in Novick and Sherman (2003) where a mean composite score of .30 was recorded for this group. Again, contrary to expectation, the self-rated skills of the participants did not predict performance in the high interactivity condition, *r*(52) = .10, *p* = .490, or in the shuffle condition, *r*(49) = .15, *p* = .285. The latency to solution also showed a non-significant relationship with experience in the high interactivity condition, *r*(52) = .07, *p* =.614 and the shuffle condition, *r*(44) = .171, *p* = .257. The proportion of time moving was not related to anagram expertise, *r*(52) = -.170, *p* =.219 nor was the rate of impasse in either the high, *r*(52) = .08, *p* =.554 or the shuffle condition, *r* (48) = -.07, *p* = .064

Summary of Quantitative Results

Only a small number of trials were excluded because people did not move the tokens and no one single participant was excluded which suggests that all participants used

73 Note that the different degrees of freedom relate to different numbers in each group after exclusions. Sixty-three participants had trials in the high conditions whereas only 59 participants had trials in the shuffle condition. Latency was only measured for those participants who solved at least problem correctly which yielded an n of 55 in the shuffle group and 63 in the high group. The groups differed as well in number for experience because of the data missing from the 8 participants. Fifty-four participants had trials in the high conditions and provided experience data and 50 in the shuffle condition. For latency, the correlations are calculated for 54 participants in the high and 46 in the shuffle.

external resources to scaffold their thinking. The high success rate and low latency to solution of the excluded trials would suggest that these were trials where the participants did not require external support to scaffold their answer as suggested by the informal observations from the pilot. Contrary to our hypothesis, participants solved more anagrams when they were allowed to interact with the tiles freely rather than when they elected to shuffle them and they were also faster to do so. While rates of impasse were the same across both conditions, participants were more likely to offer a correct solution in the high interactivity condition although the latency to solution was nor significantly different in the two groups. This indicates that when stuck it was more beneficial to move the tokens with agency that use a passive shuffle which was contrary to the hypothesis.

While participants were asked to report when they encountered impasse during the trial, if they got the anagram correct, they were also invited to report their feelings of impasse. For those trials where participants did not declare impasse during the trial, significantly more participants reported feeling stuck in the shuffle condition than the high condition but there was no different in post task impasse when participants declared impasse during the trial itself.

When it came to the time spent moving the tiles, participants were faster to engage with the tiles in the high interactivity condition but unlike in the word production in Chapter Four, the time spent interacting with the tiles was not a significant predictor of success. However, the comparison here is less informative that might first appear, the word production task was not measured in categorical success but rather with a more finely grained measure of number of words produced. Therefore, as argued above, the relationship with interactivity is likely to be qualitatively different.

Qualitative Results

Despite the theoretical justifications outlined in the introduction and the initial preliminary data in the pilot studies, there was no significant effect for the shuffle conditions and the predicted effect was, indeed, reversed. Again, in order to understand the mechanisms that allowed interactivity to scaffold problem-solving in this condition, a detailed qualitative analysis was conducted.

Nested Sample Participant Characteristics

The behavioural coding required the first author to watch each participant in the quantitative data set in close detail. This preliminary review (Heath et al., 2010) resulted in the generation of several aspects of the system which seemed salient to the whole data set. These aspects covered participant behaviour and also the role of chance. These were then assessed in greater depth through the lens of seven cases. These cases were not selected randomly (see Seawright & Gerring [2008] for a discussion on selection of cases) but rather a series of analytical notes was kept during the preliminary review which informed the opportunistic selection of cases to illustrate the observations in the preliminary round. The cases were deemed typical in terms of the observed behaviours both by myself and another member of the research team[74] who also watched the entire video corpus. All the nested sample participants were female, and the additional characteristics are displayed in Table 5.2.

---

[74] My thanks here to go Andrea Marin for her work watching the videos alongside me.

Table 5.2

Characteristics of the Nested Sample

		Verbal fluency[a]		Experience[b]	
Participant	Age	Raw	Relative[c]	Raw	Relative[c]
10	21	47	-0.82	-4.0	-1.6
32	25	74	0.78	1.1	0.3
36	20	43	-1.06	missing	missing
59	23	55	-0.35	2.0	0.6
61	21	64	0.19	1.5	0.4
62	29	85	1.44	3.5	1.2
71	22	51	-0.58	-2.0	-0.9

a Calculated as a total of the phonemic and semantic fluency. **b** Assessed using the scale from the Novick & Sherman (2003). **c** Calculated as SD from the mean of the experimental **population.**

Three main observations were generated through the preliminary review of the whole data set. First, that participants often missed environmental changes that would be helpful (missed chance), second that problem-solving in this instance advanced incrementally through a combination of heuristics and trial and error (interactivity), third that the problem solution was often dependent on an arbitrary choices made early on in the problem-solving trajectory (path dependency). This analysis was expanded on through the selected cases and additional observations were generated. This required the videos to be watched in detail both at normal speed and at a slower (70%) speed. Additionally, at key times, frame by frame analysis was required to fully reveal the micro moments in the problem-solving trajectory (Schubert, 2013). The final qualitative data set yielded over 500 annotations. An example of P10 can be found in **Appendices F** and **G**

The aim of the qualitative phase had always been to elucidate the cognitive processes in more finely grained detail but the strong effect in the opposite direction of our hypothesis made such an analysis more salient. The following analysis aims to shed light on the luck and the process of insight problem-solving more generally

Contingent Path Dependency

Problem-solving, particularly insight problem-solving is a path dependent process. By this I mean the nature of the path taken determines the likelihood of a correct solution. The initial path selected in an insight problem is necessarily done with no vision of the end process (for if the end point were clearly visible the problem would be solved analytically). Indeed, it has been suggested that people do not plan even one move ahead in some cases (Ormerod et al., 2013). The idea of path dependency echoes the ideas put forward in the Criterion of Satisfactory Progress Theory (CSPT; Ormerod et al., 2002). At each stage the problem space is restricted, and moves are selected which maximise progress towards a goal. What I would like to emphasise in these observations is that an initial choice will strongly direct the journey taken.

This seems particularly pertinent to problem-solving with anagrams. For example, it is impossible to quantify the difficulty or otherwise of an anagram (Knight & Muncer, 2011) but there are a few things that seem to make it harder. One is the similarity to a different, non - target word. Thus, if the problem solver sees a collection of letters which trigger a recognisable word (e.g. LA in MYLAD; answer MADLY), CSPT would predict that she uses that as a starting point on the problem-solving path (e.g., exploring LAD). If this bigram were unhelpful, the problem solver would be less likely to solve the problem and more likely to explore unproductive paths.

The granular data analysis suggested that this initial problem-solving move can happen arbitrarily. This can happen in the shuffle conditions where the change in array is likely to force the problem solver in a direction decided purely randomly: A fortunate tumble of the letter tiles would qualify as luck, that is random chance which directs the problem solver in the correct direction, that which points the problem solver down the wrong path would be unlucky. Take for example, P32: her third shuffle trial required her to unscramble VANIE to find NAÏVE. The first shuffle yields EIVAN which is close to EVIAN (a proper noun hence disallowed), her second VNIAE which is not dissimilar to the original layout and unlikely to be of help with no new bigrams, the final shuffle again gives her an unhelpful array - INEAV - where no letters are in the correct position. Contrast this with P59 who was aiming to unscramble MYKOS and yielded SKMOY on the second shuffle – the S and the Y anchoring the answer of SMOKY which was indeed announced at the same time as the last tile fell suggesting that it may have provided the answer for the participant. Similarly, when looking to unscramble MENGO, the same participant yielded GNEOM again on the second shuffle and again announced the answer at the same time as the last tile even after announcing impasse earlier. Thus, while neither shuffle produced the full word, it did produce an array which made the step to the full solution easy to see. However, it is notable that there were few examples of this ‘lucky’ shuffle which means that the quantitative results may be driven by a particularly unlucky set of shuffles. [75]

This path dependency, that is the importance and yet arbitrary nature of that initial choice can be seen in the high interactivity conditions as well. P10 for example (solving the anagram GNOME) creates the four letters MONE and is left with the G. Observation of other problems solvers (e.g., P32) suggests that the GM bigram can be a pivot which leads to the solution of the problem (the M and the N presumably being orthographically

[75] An additional line of research would be to somehow quantify the “luckiness” of a shuffle and present to participants in a more controlled way. However, the reflections on a lucky array in Chapter Four and Chapter Eight suggest that this would pose challenges. Perhaps the only way would be a large-scale norming project which examines performance on each array.

similar). Here P10 makes the decision to move the G to the end of the array creating MONEG rather than GMONE. This decision opens up a path away from the solution; there is a vexing contingency to this moment.

Similarly, with the same word, P36 forms the letters GNME across the tray and moves the letter O last. She places it between the G and the N, breaking a helpful bigram and frustrating her solution. Similarly, P62 is attempting to solve the anagram PROBE. She has formed the word ROBE and is left with the letter P which she places in the centre of the array, had it been placed at the start of the four letters the word would have become obvious. Take for example the same participant when faced with the word MILKY as illustrated by Figure 5.5. This sequence of actions is taken from 14 seconds into the problem-solving. The participant has spread the tiles out and is now reforming them in a line. She hovers between K and I and decides on the I which is dragged down to the M. Now MIL is across the bottom line, that is the line of interest. She now has to decide where to place the K, if she places it at the end of the MIL trigram it will spell MILK and most likely trigger a solution. Instead, she places it at the inviting gap between the I and the L – knocking the L out of place in the process, a move which she reinforces by tidying it further away. That decision means that she is now looking at a tempting but unhelpful trigram of MIK rather than the more useful MILK.

It is important to note that these decisions are made quickly, and do not appear to be mediated by strategic thinking. The series of movements illustrated in Figure 5.5 took place over the course of only 4 seconds. Equally, the effect of either luckiness or unluckiness is only able to be assessed by an observer who already knows the outcome[76].

[76] I will discuss this at greater length in Chapter Eight.

However, the arbitrary nature of that choice first supports theories that suggest a lack of look ahead in problem-solving and second, questions models that attribute all epistemic credit to the person when the moves can themselves be contingent on nothing more than a lucky hunch. It is easy to trace moments when unplanned decisions lead to unfortunate pathways, it is less easy to distinguish between equally frivolous decisions which lead to a successful outcome and those which are the result of sober planning.

Microserendipity.

There were several moments of microserendipity, that is when moves by the problem solver trigger unexpected configurations in the array which are then noticed and used to direct the problem-solving trajectory. For example, P10 when solving MILKY first spells out the word MILK and once this is formed, she notices the left-over Y and uses it to create the word MILKY. Her voice rises in pitch in excitement and the move which attaches the Y is faster than the previous slightly laboured spelling out. The solution is triggered by a change in the array rather than a preformed plan. P32 has a similar moment when the solution – the event pivot – is triggered by an unplanned change in the array as demonstrated by Figure 5.6. Here the participant is solving the anagram GNOME. She moves the G and the E out of the line (moving the middle tiles is a common strategy) with the G going above the array and the E going below, she then moves to close the space between the M and the O which means the M is directly lined under the G. This GM bigram appears to lead to the rapid solution, tested first and announced 6 and a half seconds later.

While I could outline further examples, the most striking thing revealed by the granular analysis was that the participants often missed the chance movements created by

Figure 5.6

P62: A Demonstration of the Contingency of Solution Choices.In this Case If P62 Had Moved the K Down to Line up With the End of MIL, She Would Have Formed MILK. Instead She Forms the Trigam MIK.

Figure 5.7

P32: A Moment of Microserendipity. In this Case Fiddling With The G and E Led To The Array Where the G and the M Are Arranged Such That the Word GNOME is Triggered

the shifting environment (see Figure 5.7 for exemplars). In the words of P71: “I feel like I can see it but I can’t”. This aspect of participant behaviour has already been noticed in passing in other qualitative analysis of problem-solving: Steffensen et al. (2016) narrate how the participant in their task solves the problem after an accidental overlap (a necessary step towards solving the 17 animals problem), yet this overlap had accidentally occurred twice previously but was ignored and dismantled. This missed chance is the same as that in Chapter Four. This lack of noticing becomes more salient in a task such an anagram task in which there is only one answer and the cost of ignoring the correct answer is greater than the benefit of quick moves in generating many choices.

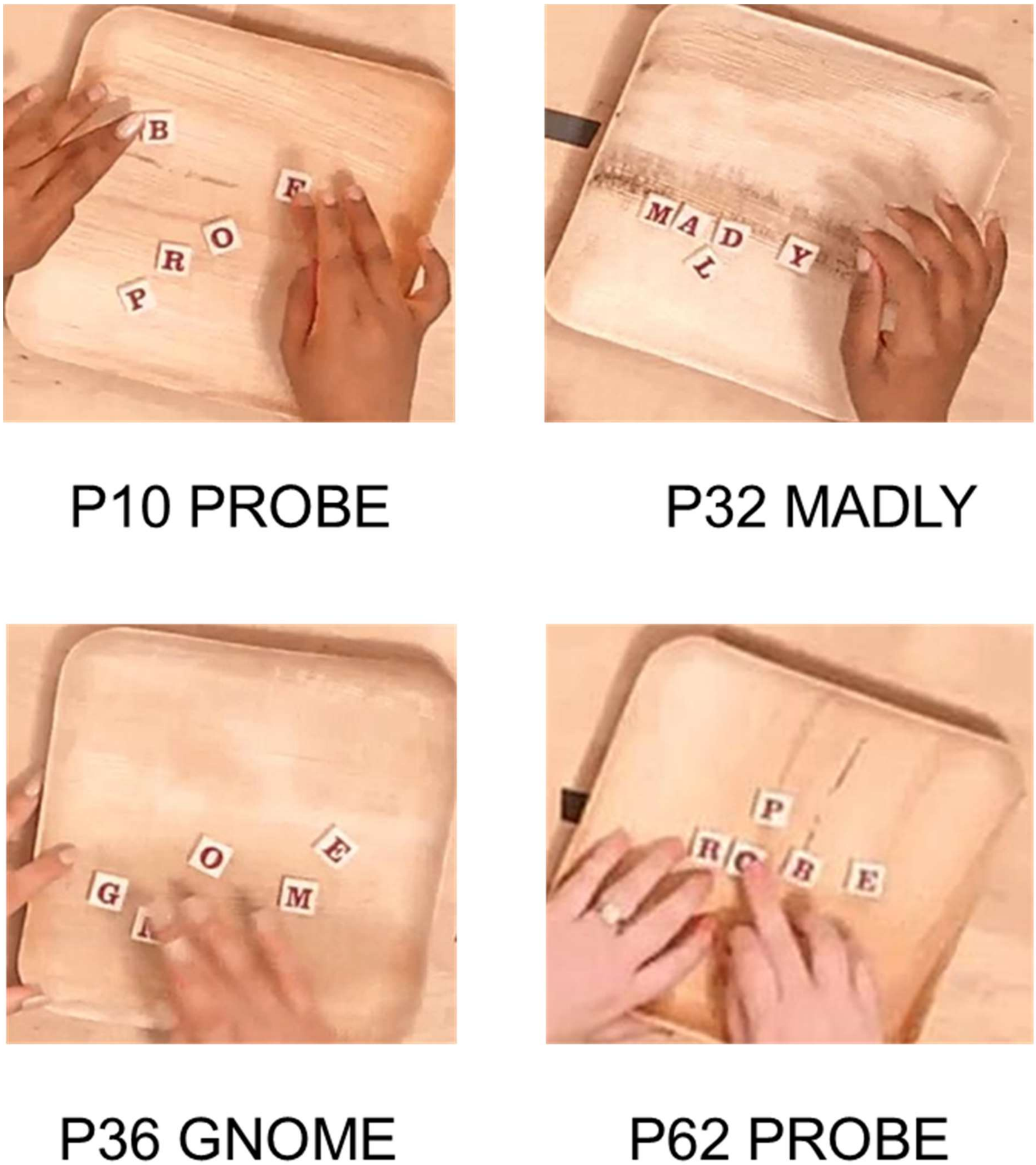
Impasse

The traditional model of insight problem-solving requires an impasse. Indeed, Ohlsson writes, “Without the impasse, there is no insight, only smooth progress’ (1992, p. 4). Participants labour away on a problem until they realise that their initial problem representation is incorrect and then they meet a moment of impasse before there is a helpful or unhelpful restructuring and so the cycle continues. However, the requirement for impasse is no longer central to theory of insight, for example, Danek labels it as optional in her model of non-monotonic problem-solving (Danek, 2018; Chapter Three). Furthermore, Fleck and Weisberg (2013) report low levels of impasse using verbal protocols during problem-solving with traditional insight problems.

Impasse is currently measured in inconsistent ways: through participant report (Fedor et al., 2015) or through assumed behaviours (Öllinger et al., 2014). In my study, I asked people to report their moments of impasse, however the behaviours do not mirror the self -reports. Take P10 being asked to find the word MINUS from the initial arrangement of SINUM. For the first 8 seconds of the trial she speaks to herself and rearranges the tiles

Figure 5.8

Moments of Missed Serendipity



in an aimless fashion. After 13 seconds, she announces that the answer is MUSIC and despite being told that this is incorrect, she forms the word MUSI and is disappointed to see there is no C. She continues to play with the tiles in an aimless fashion, making the start of MUSI several times, muttering until she declares impasse after 74 seconds. There is a tension here in how we understand impasse.

By any definition the behaviours displayed by the participant here indicate that she is encountering impasse – she offers the wrong answer, she continues to return to an unhelpful representation, her behaviour indicates that she has no coherent plan–but she does not declare impasse as we would understand it. It is also noticeable that impasse does not appear to elicit further problem-solving attempts in all the participants and in fact appears to be a proxy for giving up. For example, P62 after 44 seconds attempting to solve the anagram SMOKY announces "I'm stuck on this one" and physically pushes the tray away. Soon after announcing impasse, P10 picks the tray up and tilts it so that the tiles fall randomly indicating that she has completely given up.

Material Traces

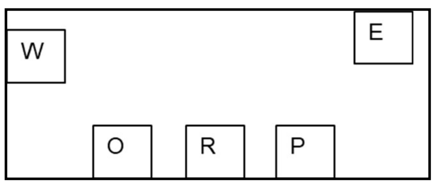
When problem-solving takes place in an environment low in interactivity, the cognitive processes can only be theorised from proxy measures such as response time (although response times are not always measured, e.g., Chuderski et al, 2020). Problem-solving in a high interactivity environment allows us to trace the problem-solving act as it happens over time and the use of lettered tiles in this instance makes this process clear. Similar to Fleck and Weisberg (2013), I found diverse methods of solving the anagrams suggesting that even relatively similar stimuli cannot be relied upon to elicit the same response. Some of the trials were solved directly–a pop out solution as described by

Novick and Sherman (2003) – other after consideration and yet others solved piecemeal with moments of microserendipity crossing with purposeful moves.

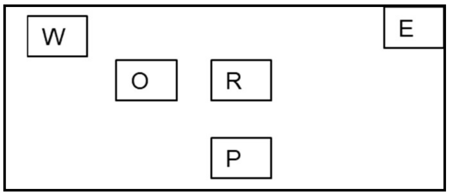
Take for example P71 solving the anagram POWER. The 10 seconds before a correct solution was announced are described in detail in Table 5.3. As can be clearly seen, the answer unfolds through a reciprocal moulding of the array; P71's gradual realisation of the answer directs the moves she makes which further support her understanding of the answer. While the pivotal event (at 19.779) seems to be the moving of the W tile which knocks into the R (a further example of microserendipity) and pulls it with it to create the useful trigger this is not the only important action. Instead, we can see a gradual unfolding of understanding over the course of the 10 seconds facilitated by purposeful moves, moments of consideration and fiddling in coordination with the accidental move of the R. Even the understanding that the word is POWER requires the participant to check before she makes the announcement, the mental representation of the word not secure until it is also represented externally, the realisation in the mind contingent on the realisation in the world.

Table 5.3

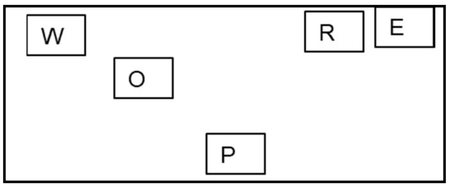
The Process of Solving the Anagram POWER

Time (s:ms from start of problem)	Description	Resulting Array
15.241	Starts to move the P	

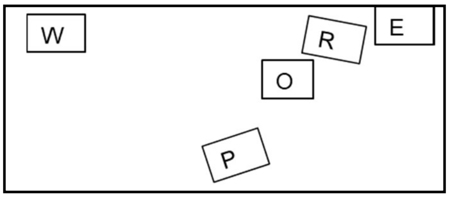
15.741 P dragged to below the main array



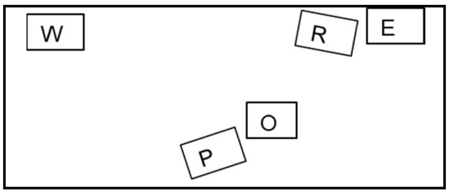
16.375 R moved out of way



17.310 O moved down next to P



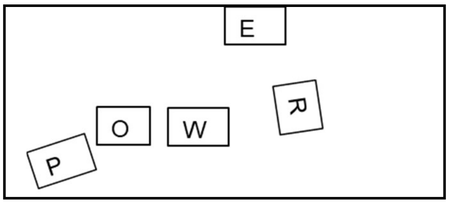
18.411 P and O grouped together



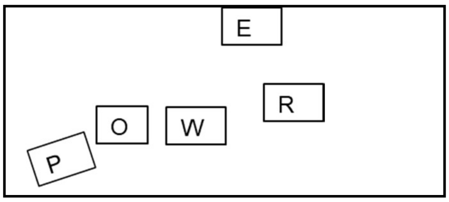
18.411-19.178 Array is considered, head moved back, fingers off tiles

19.178 The letter W is selected for the next move

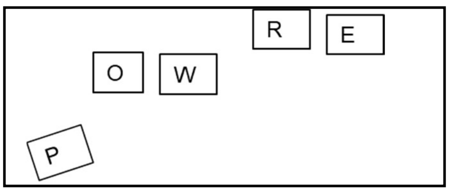
19.779 W is moved next to P and O. the move knocks the R tile and pulls it with the W tile in line with the P and the O but tilted on its side



20.646 P71 tidies the array so that R is straight



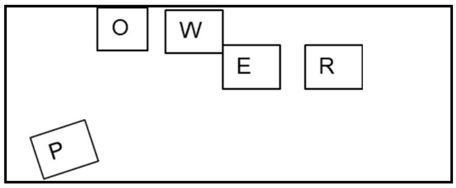
21.780 P71 moves the E down so the tiles are in a straight line.



21.780 - 23.180 P71 considers the array

23.180 P71 traces her fingers over the R and the E

24.450 P71 forms the word POWER



25.404 P71 announces the correct answer

Discussion

Problem-solving in rich and varied environments is a complex mix of internal factors, external intrusions and conscious and non-conscious couplings of the two. Selecting which aspects of the dynamic interplay of people and things to investigate in the laboratory is not an easy task. Problem-solving will always be a richly contextual activity and adding the variable of random environmental fluctuations further complicates matters. However, taking the material and external world into consideration is important if we hold to the theoretical perspective that moves over objects in the world are constitutive of cognition.

First, while the extent to which interactivity can augment cognitive performance and the implications for the ontological locus of cognition is not an uninteresting or unimportant question, the behavioural analysis points to something more fundamental. The first hypothesis that people would interact with the world when given the opportunity was upheld. Every participant opted to make use of the material world in at least one trial. Whether cognition does or does not *require* the external world is less important than noting that it frequently involves it[77]. Taking aggregate measures of what *can* be done is possibly not as useful as assessing what *is* done if we wish to understand situated problem-solving. The evidence that every participant made use of the scaffolding afforded by the lettered tiles on at least one trial suggests that whether or not it has an augmentative effect, problem-solving that ignores the role of the environment is a poor shadow of how people think.

However, the efficiency with which participants interacted with the environment varied. For one, there was no relationship between a correct solution and the time spent moving the tokens unlike in Chapter Four. This suggests that environmental engagement is not a panacea and nor is it consistently and linearly augmentative in a one solution task of this type. While it will support strategy and incremental change in a task with a continuous outcome variable, there is a more complex and contingent relationship between moving and success in a task of this type. It may be that it functions as a form of affective and motivational support which is why the feelings of impasse were significantly lower in the high than the shuffle condition while the declarations were not. This was an unexpected disconnect and requires further investigation. In addition, there was a curious lack of attention to helpful changes in the array. The notion that the environment can yield a solution and that this is unproblematically recognised and taken up would appear to be naïve. The act of noticing appears to be a crucial component of the interaction with chance and deserves further investigation.

[77] As Clark (2010, p. 24) puts it: "(…) we often do lots of stuff (dreaming, planning, musing) entirely in our heads, using inner surrogates for absent states of affairs. But it is surely worth noticing (...) just how much of our cognitive activity is *not* like that (…) it often seems that brains like ours (though there is considerable room for individual differences in these regards) will go to extraordinary lengths to *avoid* having to resort to (…) fully environmentally detached reflection (…)" (emphasis in the original).

Luck and Serendipity

Given the anecdotal evidence about the role of luck and accident in problem-solving, the current study was designed to test how important random material changes were to the array. The quantitative phase of the current study tested the hypothesis that luck is enough to drive problem-solving. It was hypothesised that a task such as an anagram would be particularly likely to be supported by a random generation of hints. The data here suggest that the opposite is true: Interacting with the tokens led to a higher overall solution rate and faster solutions than luck and mental effort alone.

There is an ongoing debate in the field of luck and serendipity over the extent to which epistemic credit should be awarded for things attributable to accident (Arfini et al., 2018; Sand, 2020). The overall evidence from the experimental manipulation was that luck is not enough to spark problem-solving, rather an active interaction and implementation of that luck is required: Chance has to be enacted. This has implications for the growing field of chance discovery. However, it is worth noting that at the highly granular level there seems to be an unclear division between luck and serendipity in this task. No one magicked the correct answer in a shuffle. Indeed, the number of missed serendipity moments suggest that even if the world throws an answer it is not guaranteed that this will be noticed by the problem solver and become significant. Therefore, by mapping the moments when chance is not enacted, our data suggest it is unclear to what extent pure, non-agentic luck is even possible. If it is necessary for a concept of luck to bear some significance to the problem solver, it may be that the act of bestowing significance on an event involves a form of agency. Further theoretical and empirical refinements become necessary.

The highly granular analysis provided additional evidence for idiosyncratic problem-solving methods both between and within subjects. This observation supports previous evidence from other more detailed analysis (Bilalić et al., 2019; Cushen & Wiley, 2012; Steffensen et al., 2016). Participants solved some anagrams quickly with little obvious effort, some slowly and incrementally and some not at all. This suggests that simply aggregating data across participants may hide important patterns. The growing evidence for diverse methods in problem-solving is further complicated in insight research by the number of different categories of problems (Batchelder & Alexander, 2012). This suggests that large scale studies which both assume the problem-solving process from the problem and collapse across diverse problem types may only provide a limited view of the complexity of insight problem-solving.

Impasse

Ohlsson (1992, p. 4) defines impasse as “a mental state in which all problem-solving has come to a halt” and goes on to suggest “behaviourally, impasses are characterised by the cessation of problem-solving activity. The triangulation of data here taken from mid task self-report, post-task self-report and informal observation of video data here indicates two things: First, that the linear model of impasse followed by restructuring also hinges on motivation which supports the data from Fedor et al. (2015) who found a significant effect of success on impasse–those who declared impasse were much more likely to fail. A declaration of impasse at times was tantamount to giving up. Second, that impasse may not be consciously experienced by an experimental participant while in the ‘flow’ of problem-solving. It is also worth noting in this respect that while there was a significant difference between the conditions in post task feelings of insight this did not replicate in the self-reporting of impasse during the task suggesting that each are measuring different phenomenon. Externalising cognitive processes appears to support the affective dimension of problem-solving (Bilda et al., 2006; Guthrie & Vallée-Tourangeau, 2015). Perhaps the slowness to declare impasse in a materially rich situation is because the very material situation does not invite it–the menu of problem-solving options is far richer and more inviting in a high or even a shuffle condition rather a low interactivity environment–or perhaps it was also because the very definition of impasse requires the participant to give up.

Mixed Methods Research

From the position of a research program underwritten by a commitment to interactivity – i.e., problem-solving as a process that unfolds in the coordination of people and things – then the particular model of mixed methods research followed here provides a granular insight into the nature of that coupling and the manner in which the reciprocal, looping process of thought unfolds. Such a method also supports a more detailed understanding of insight problem-solving alongside the use of eye tracking measures (Bilalić et al., 2019) and think aloud protocols (Fleck & Weisberg, 2013). The evidence presented here for the uncertain status of impasse, for example, would only be possible through a combination of self-report and behavioural observation.

It is also important as the research field in serendipity and chance discovery seeks empirical evidence to support its work on how we interact with chance and uncertainty. The methods employed here allow us to get closer to assessing how lucky participants are, albeit in a toy task and in the controlled environment of the psychologist's laboratory. By demarcating temporal boundaries and providing clear binary outcomes and the use of identifiable way markers (the different letters) that trace the route to solution, I suggest that it is possible to trace both luck and serendipity at the level of micro movements. How this will scale up remains to be seen.

Chapter Six :A Grounded Theory Analysis of First Order Problem-solving

Chapters Four and Five introduced the benefits of detailed behavioural coding under the framework of kinenoetic analysis to understand problem strategy in a well instrumentalised task environment. The data were analysed in three broad ways: quantitively, through the application of preconceived codes which turn qualitative data into quantitative codes and through cognitive case studies. The research reported in this chapter takes a more qualitative turn. Originally conceived as an extension of research on the mechanisms of performance in high interactivity research, the unanticipated and theoretically puzzling "failure" of the experiment generated the opportunity for inductive, qualitative analysis to generate novel substantive theories relating to first order problem-solving research and formal theories relating to experimental research more broadly. As these observations arose from the data, they were analysed under the framework of Grounded Theory (Glaser & Strauss, 1967) although there were several key deviations from strict Grounded Theory Method which are outlined below

In short, the study reported here aims to understand a system where the existing theoretical frameworks break down and to offer some explanations for this grounded in the video data collected during the study. I report the quantitative data first before moving from the hypothetico-deductive approach implicit in the analysis to an inductive analysis which generates two broader formal explanatory frameworks: Environmental information can slow down cognitive function as well as augment it and cognitive systems extend beyond the behaviour of interest to encompass the laboratory set up which is saturated with opportunities for interactivity.

As outlined in Chapter One, the dualism behind early incarnations of the extended mind thesis assumes two things: That methodological solipsism is achievable (albeit although unhelpful when it comes to understanding “real world problem-solving”) and that interactivity always has a scaffolding effect (see also in this regard the case studies selected in Chapter Four). This latter led to experimental designs which contrasted high and low interactivity, if performance were better in a situation in which the participant were given a movable and interactive display then that established claims for the augmentative effect of interactivity and thus its existence as a phenomenon of interest (see Chapters Two and Three).

The augmentative effect of interactivity has been broadly demonstrated by the first two studies reported in Chapters Four and Five. Although there was no difference between conditions in the main study in Chapter Four, the detailed analysis of behaviour revealed a strong positive correlation between the time spent engaging with the movable interface and the number of words produced. The interactive condition also supported participants to enact a more efficient strategy. However, the case study data did demonstrate that interactivity (in the form of movable problem representations) is not a universal panacea (see also F. Vallée-Tourangeau et al, 2020). Chapter Five discussed the difficulty of contrasting experimental conditions based on the behaviour in each and so the main study simply contrasted two different types of interactivity and found that an active coupling with the world produced superior performance to random changes in the environment. However, again the detailed case studies revealed that the relationship between agent and world is complex and no guarantee of success.

The Omniscient Problem Solver

The idiosyncratic paths revealed by the granular analysis of individual problem solvers suggests that notion of an agent centric cognitive system in which the brain bound thinker selects the correct tools to systematically augment cognition is problematic once we step outside well-structured problems. For interactivity to be consistently beneficial requires an omniscient agent who selects the correct resources to solve the problem with which they are presented. Problem-solving is broadly the movement from a state of ignorance to a state of knowledge (Arfini, 2019). There are two forms this ignorance can take, ignorance of the process of problem solution or ignorance of the answer. The most efficient pathway to solution can only be directly and easily enacted when the solution is already known and can be traced backwards without the risk of a false start or a complete divergence from the path. This requires fore knowledge of the correct answer.

With analytical problems while the answer is unknown, the process of getting to the answer is clear and known. For example, with something such as mental arithmetic, while the problem solver may not know the solution, they will know the most efficient process and the simple operators to evince it[78]. These steps will have been already predetermined culturally and through personal experience. Therefore, it is likely that the agent will select the correct objects and the correct actions over those objects. This means that an agent-centric model of interactivity should yield supportive empirical data, however it is unclear how far such a model would extend.

For some insight problems, the answer is given but it is the method of reaching the

[78] Indeed, the accuracy of the solution is not clear in itself but is predicated on an accurate following of the steps. We only trust an answer to a mental arithmetic problem not because of the answer itself but because we trust in the steps that led to the answer.

answer which is unclear. Take for example, the triangle of coins problem (e.g. F. Vallée-Tourangeau et al., 2020) which requires participants to move from one triangle shape to another in a restricted number of moves. The end state is clear and easily recognisable. To some extent this is the same with anagrams where the target word can be recognised by the problem solver once it is revealed. Thus it is not surprising that Chapters Four and Five report data which suggests that participants can accidentally generate a word (see also F.Vallée-Tourangeau et al., 2020 for video data of a participant recognizing the answer to the triangle of coins problem[79]). In this instance the process becomes unimportant, the answer is easily recognised as being correct and functions as its own accuracy check. You can "see" that you are right. The importance of the recognition of the correct answer becomes clear if we look at the issues with the pilot trials for Chapter Five (**Appendix D**). Over the course of these trials the participants who were not familiar with the target word had no measure of accuracy for the correct answer and so could not recognise it.

However, in many ill-structured problems or non-analytical problems, the problem solver is in double ignorance: They do not know the correct solution or the manner of approaching the correct solution. In terms of extended mind theories, this means that they will not necessarily select the correct objects or actions over them but merely those that satisfy the needs at the time (echoing the Criterion of Satisfactory Progress Theory; MacGregor et al., 2001; Ormerod et al., 2013). As demonstrated in Chapters Four and Five, incorrect initial representations can be reified and slow down processing leading to a detrimental effect. Equally, the benefit of reifying the answer is somewhat dissipated if the

[79] https://osf.io/7nbkt/?view_only=4a3fe2db192f4dbabca859f7ce0e70e2, note at 0:00:07 the double take expressed with both hands splayed open when the participant created the solution.

participant does not recognise it as a solution[80]. This does not mean that interactivity has "failed" but rather that the system (here defined in terms of the problem solver and the movable objects) is performing less efficiently according to the parameters of the task[81].

The problems presented in the Chapters Four and Five showed a significant benefit from interactivity, that is engaging with the problem space. For the word production task, this was in part because interactivity scaffolded the already existing knowledge of the correct pathway to solution. Those participants who already knew the best way to perform performed better across all conditions, but the optimum strategy was easier to enact in this condition. This suggests that interactivity functioned primarily to scaffold a form of "thoughtful" thinking. However, it seems different problem types are likely to be affected differently by interactivity because they require different and more opaque cognitive processes. When there is not an obvious problem-solving process to scaffold then the role of interactivity will be different, and it may even have a detrimental effect. Systemic thinking assumes the soft assembly of dynamic systems which will form when required to undertake a cognitive task but the efficacy or not of those system is not a predetermined outcome (F. Vallée-Tourangeau et al., 2020).

Such a claim leaves experimental research in interactivity in a difficult place. The first two empirical chapters of this thesis have started to work through the impact of shifting from an experimental manipulation focused on outcomes to a process-based

[80] To return to the moments of serendipity that were outlined in Chapters Four and Five, this "act of noticing" (Rubin et al., 2011) is considered an important part of the connection between environmental chance and the actions of the agent. The data I discuss above also suggests that the mind and world may not be automatically, consistently and unproblematically connected, rather that there has to be a spark of some sort.

[81] It may be that it is performing more efficiently in terms of other task parameters less often tested. For example, there is preliminary evidence for importance of interactivity on flow and affect (Bilda et al., 2006; Guthrie & Vallée-Tourangeau, 2015) and I certainly informally witnessed that when carrying out research on children. In this instance interactivity may augment affective state even when it does not affect performance but this would not be captured by a pure performance based approach.

account. Chapter Four described a word production task which demonstrated that it is not enough to simply allocate participants to a high interactivity condition, the behaviour in those conditions needs to be monitored. An analysis based simply on the performance data would have provided a poor explanation of what actually happened throughout the course of the environment. The justification for the subsequent experimental design laid out in Chapter Five also queries the utility of a straight comparison design and suggests that the focus of interactivity research should rather be how cognitive ecosystems form. This has led to interactivity becoming both an experimental manipulation and a method of tracking thought through the movement of objects: The method of kinenoetic analysis.

The Problem of Insight Problems

The class of problems described as insight problems has already been discussed in Chapters Three and Five. However, what has perhaps been less well emphasized is the idiosyncratic nature of the problems in this group (Batchelder & Alexander, 2012). This was resolved to some extent by the use of compound remote associates and even anagrams in studies which require multiple trials for validity (see Footnote 66,p.149) but it still remains a problem because inferences drawn from one problem do not easily map onto those drawn from another problem. For example, the cognitive mechanisms underlying the solution of an insight problem framed as a riddle are likely to be qualitatively different to one framed as a visuo-spatial problem.

Beyond differences across problem types, it appears that the same problems can yield different levels of the affective experience of "insight" across different people, suggesting different methods of reaching the same solution. This leads those who are interested in the phenomenological markers of insight to suggest that "insight is not in the problem" (Webb et al., 2016), It also appears from preliminary data that levels of insight it may even constitute an individual difference (Webb et al., 2019) rather than an indication of problem solution or process. Thus the argument for a false "aha" in Danek's (2018) model discussed in Chapter Two. If this is the case, then there is a danger in relying on reports of affective experience as a proxy for cognitive processes (see also Chapter Five; cf Danek et al, 2020).

The Sock Problem

The "sock problem" is a classic insight problem. The participant is asked "If you have black socks and brown socks in a drawer mixed in a ratio of 4 to 5, how many socks will you have to take out to make sure that you have a pair of the same colour?". The problem masquerades as a mathematical problem – the conversational pragmatics forefronts the 4:5– but it is misleading information. The answer is three no matter what the ratio. It is possible to pull a pair with the first two socks but also to pull a brown and a black, however the third sock would necessarily have to match with one or the other. The ratio of brown socks to black socks is immaterial and a distraction designed to set the participant along the wrong path.

Bowden et al (2005, p. 323) suggest that this is only a problem if you approach it mathematically but if you use a 'What if' strategy then you are far more likely to simply solve the problem: "That is, if the solver asks, 'What if I take out a black sock then a brown sock? I would only need one more sock of either color to have a pair of the same color.' No insight is required." Jones (2003, p. 12) suggests that: "The insight here involves moving from a representation of the problem based on mathematics to a representation of the problem based on imagining yourself removing the socks." That this strategy is the more efficient one is suggested by Chu and MacGregor (2011, p. 211) who suggest "The problem solver may only have to run through a couple of trials in her head before arriving at the conclusion that, at most, you only need to draw three socks to match a pair". In a later description of Fleck and Weisberg (2013), Weisberg (2015, p. 32) noticed that "some participants solved the Socks problem on their first attempt, by imagining taking the socks out of the drawer, and considering the information available at each step."). Such a mental simulation strategy could be replaced with actual behaviour in a first order problem-solving environment.

In other words, the problem solution invites an initial mathematical representation but requires the participant to not only see the problem as not requiring a mathematically based solution but also to see themselves in a real world environment and solving it in a trial and error way. Therefore, it seems almost trivial to suggest that a high interactivity condition would augment problem-solving, rather than the problem solver imagining herself drawing socks from a bag, she could actually do that. Indeed, this is what is predicted by F. Vallée-Tourangeau and March (2019, p. 3)

> Let us transpose the socks problem, as described above, from verbal riddle into a physical model of the problem (from second- to first-order). Imagine a duffle bag with 40 black socks and 50 brown socks. Our participant reads the problem description and is invited to determine how many socks she will have to sample before getting a pair of matching color. She is told she can dig into the bag and pull a few socks, one at a time, to help her solve the problem. The misleading ratio information in the problem description might not attract her attention as much as it would otherwise were she only presented with the riddle without a physical model of the problem. She might not know how to solve the problem; she starts pulling a few individual socks, not strategically, not with a plan in mind, but simply exploring, interacting with the problem and observing results. The misleading ratio information quickly fails to exert any attraction; rather she's looking at the results of her sampling from the bag. She may pull two black socks from the start, tempted to say that the answer might simply be ''two'', but realises that she's been lucky, pulls a third one and fourth one, and the solution dawns on her; the solution is

> distilled through action and results. The insight, if there is one to experience, takes place when she observes the results of her actions.

This assertion is not tested by the authors – it functions simply as an "intuition pump", however it is justified both theoretically and in light of the prior theorising on the routes to solution of this particular problem.

Intuition Pumps

There is no standard definition for a thought experiment, and they take on many forms (Stuart, 2020). While they have been a feature of philosophy and, arguably scientific enquiry, they are not without controversy most notably in terms of the "relatively simple epistemological challenge" (Brown & Fehige, 2019) that is posed when thought experiments are used to derive new knowledge about reality. At the heart of this challenge is what has been called the paradox of the thought experiment, which is essentially twofold: How can knowledge about the real world be generated without reference to that real world and how can new knowledge be generated from a closed mental system? In this way, it is similar to the argument for the paucity of an internal model of unstructured problem-solving.[82] If problem-solving requires both new knowledge and contact with the world to enact that knowledge, then so too do thought experiments at which point they cease to be mental simulations.

Thagard (2014) argues that while thought experiments can play a role in scientific

[82] And indeed of creativity. This central paradox echoes Ohlsson (2018, p. 12) on creativity: "The main puzzle of creative cognition is that it can produce novel concepts, beliefs, problem solutions and products that are not in anyone's prior experiences. How is this possible? Where does the novelty come from?". In Ross & Arfini (forthcoming), we make the argument that this puzzle stems from the dogged adherence to an internalist model of creativity. A closed system can spontaneously rearrange to generate novelty, but such moments are unusual, the puzzle is far more simply solved by opening up the system.

discovery they can also be harmful. Intuition pumps are a form of thought experiment which Thagard cites as particularly problematic because of this reliance on intuition which he suggests is “poorly suited to provide evidence for beliefs” (p.289). Thagard characterises these cognitive tools as stories which generate intuitive judgement about the situation at hand. That is, they require the reader to run through the situation mentally and simulate possible options, but they do not have the rigour of a philosophical argument; what is right or not is generated from intuitions which “are not much better than those of astrologers” (p.293). Dennett (who coined the term “intuition pump”) agrees with Thagard on the lack of rigour but views them in part as a rhetorical device: “little stories designed to provoke a heartfelt, table thumping intuition” (p.6) which is certainly the case in this instance—the gentle well-formed story of the problem solver as laid out by F. Vallée-Tourangeau and March has an intuitive logic and extends only slightly other claims about this problem. Dennett goes on to suggest that these cognitive tools are necessary for understanding some of the big topics that are not amenable to empirical examination. This seems a reasonable proposition when it comes to ideal types which are abstracted from material reality and therefore cannot be investigated empirically, it becomes less convincing when dealing with things which are not referencing these ideal types.

Of the seven sins of thought experiments which Thagard raises, it is perhaps his highlighting of the under specification of typical thought experiments which is useful in considering the transfer from mental to real world. We have seen that the ability of the mind to simulate is restricted in scope and, in thought experiments, this owes much to this under specification and focus on ideal types. In other words, a mental simulation works with the information that is required to solve the problem but the information which is required to solve the problem is necessarily prespecified in a thought experiment by what is already in the system. This again requires an omniscient thinker. The search for ideal types and abstracted situations ignores that it is not always clear what the underlying salient conditions are.

It has been a central argument of this thesis that mental simulations are not enough and that interactions with the environment can generate novel thoughts either through scaffolding existing thinking processes to reach beyond the cognitive capacities of the agent or by introducing novel information into that system. In this instance it becomes particularly important that we do not rely on thought experiments to demonstrate the proof of hypotheses which can be established through empirical means. Simply put, there is no need for a thought experiment when we can view the results of an actual experiment. Thought experiments are only necessary when they consider things which cannot be attended to empirically.

The thought experiment outlined above introduces a novel mechanism through which interactivity is beneficial which take us beyond an agent centric approach – that of unplanned and aimless fiddling. The sock problem does not require cognitive scaffolding, nor does it require the introduction of novel information to the system, rather it requires the repression of information. Any benefits yielded by interactivity should result from the shifting of the problem presentation through action and the transfer into a different problem space, away from an arithmetic problem towards a more mundane activity. The logic and the mathematics involved are not difficult, so the mechanisms at work in an interactivity boost can only be those of representation change and triggering a sudden insight (outsight[83]) through action. This means that the evidence that could be generated by

[83] Vallée-Tourangeau and March argue that when an engaged thinker sees something in the environment which triggers this feeling of insight that this is better classed as "outsight".

testing this would strengthen arguments that interactivity is more beneficial than simply acting as a scaffolding extension. As F. Vallée-Tourangeau and March argue, the benefit here is that the answer unfolds in front of the problem solver.

In this instance, the problem-solving described above in the thought experiment relies on a series of exaptative gestures (Ross, forthcoming a, b). Exaptation describes the reuse of an existing artifact for a new purpose. It has been suggested that exaptation is an important driver of innovation because it allows for unplanned novelty to arise (Andriani et al., 2017). I borrow the term to refer to actions which have no initial purpose (fiddling for example) and from which purpose emerges once the results are seen. Exaptative actions are different from epistemic or pragmatic actions (Kirsh & Maglio, 1994; see Chapter 2) because they decentre the human agent—she proceeds "not with a plan in mind"—and their purpose changes through action.

It is not implausible that such exaptative actions would be likely to be important in insight problem-solving in a first order environment. Take the example of the sock problem, scaffolding or extending the mental workspace is not necessary or helpful. An understanding of the problem would need no external scaffold. Similarly the solution of the 17 animals problem reported in Steffensen et al. (2016) occurred when the participant was engaged in a form of aesthetic tidying, the original purpose of the action was not to generate an answer but the gesture became an answer- generating gesture. Such actions over objects are very different to the actions over objects normally attended to in interactivity research[84] which suggest the actions act as a deliberate scaffold. Therefore, the

[84] Indeed, my own research reported in Chapter Four excluded these gestures from analysis focusing only on "meaningful" moves. In Chapter Eight, I discuss this tension between a meaningful and non-meaningful move.

claim made in the thought experiment is perhaps stronger than the authors may realise. However, it also reflects the moments of microserendipity identified in Chapters Four and Five and so deserves further investigation. [85]

Thought experiments are necessary in hypothesis generation but when they are moving beyond ideal types and conceptual analysis to firm empirical predictions, they should be tested to ensure the underlying premises are sound. Chapters Four and Five have already demonstrated that the mechanisms through which the initial hypotheses were formed do not map onto the data reporting what actually happened. Thought experiments have a validity when they are dealing with things which cannot be tested empirically but they are less valid when they are dealing with things which are situated in a real world. Therefore, given the theoretical import of the claim, it deserves to be tested.

The Current Study

The intuition pump outlined in F. Vallée-Tourangeau and March (2019) makes two predictions. First, that the use of an actual bag of socks will lead to augmented performance but second it outlines the mechanisms through which this performance would be augmented, specifically two ways: First, that the real-world situation would lead the participant to forget about the misleading ratio information and second, that the participant would be able to view the answer as she pulled socks out of the bag. This clear behavioural, process hypothesis has a firm theoretical basis as outlined above. The current study was designed to test the behavioural hypothesis alongside the performance hypothesis.

Participants were therefore invited to solve the problem in both the high and the

---

[85] These exaptative gestures become particularly important in Chapter Seven where they give rise to the form of the artwork explored.

low interactivity condition, but the focus of the analysis was the behaviour that they displayed. Chapters Four and Five report studies which are quantitative dominant with the qualitative phase functioning as an explanatory phase: the structure of a kinenoetic analysis. The initial study reported here was intended to follow a similar analysis plan but diverged as it became clear that there was a failure of the experimental manipulation. I therefore present the quantitative method and results before moving to the qualitative phase which yielded more substantive conclusions.

There were therefore two formal hypotheses: The first, a performance hypothesis: that a high interactivity environment would augment interactivity. The second, a process hypothesis: that high interactivity would lead participants to (a) disregard the ratio (b) generate moments of microserendipity through observing and recognising the answer in their actions.

Quantitative Section

Method

Participants[86]

Forty-one participants (a mixture of undergraduate and post graduate psychology students) were recruited in return for course credits. One participant was excluded for knowing the problem, one for not adhering to the stipulations of having English as a first

[86] Although the primary aim of this study was qualitative coding of strategy from the start using KA to assess the process hypothesis and there is always a balance to be struck between participant numbers and granularity of analysis, my initial target participant number was 60 to produce a sample size adequately power to test the performance hypothesis. However, data collection had to be stopped owing to Covid-19 and the ban on human psychological testing of this kind. The research described here would not be able to be moved online nor will it be able to be conducted while social distancing measures are in place because of the risk involved with the use of material objects. Therefore, the decision was made to analyse the quantitative data although it was still underpowered. The subsequent analysis suggests that the quantitative results would have remained broadly the same.

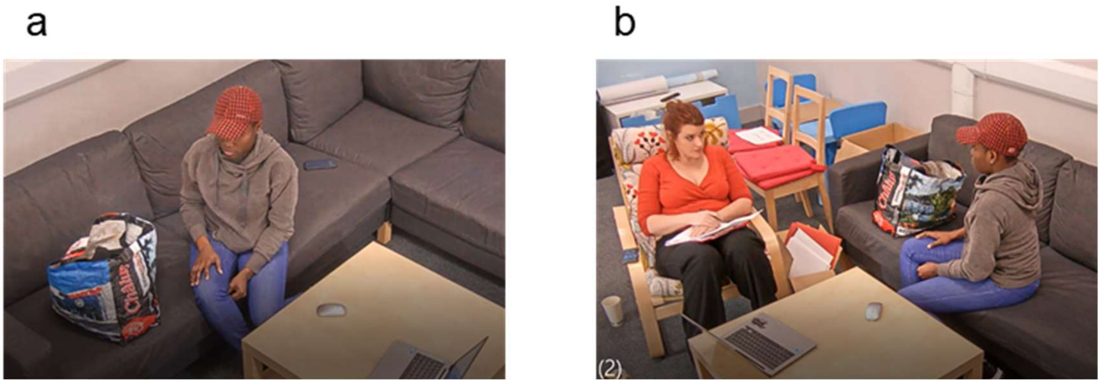
language and one because of data loss. This left 38 (*F*=31) participants with a mean age of 26.28 years (*SD* = 12.14). The participants were assigned in turn to each condition. Eighteen participants took part in the low interactivity condition and 20 participants took part in the high interactivity condition.

Procedure

Participants were filmed in a purpose built lab with three in-built cameras. To minimise the differences between the conditions both high and low interactivity conditions had the same initial set up: The participant was presented with a bag of 90 socks, 50 were black and 40 were brown in both conditions and the only difference was that the participants were allowed to pull socks from the bag in the high interactivity condition (see Figure 6.1).

Figure 6.1

The Experimental Setup. Participants Were Given a Bag With 50 Black Socks and 40 Brown Socks (panel a). The Researcher Stayed in the Room to Give Instructions and Make Notes (panel b; Screen Shots Taken At 1:09)



The original instructions were taken from the existing literature and ran thus: "If you have a drawer with brown socks and black socks mixed in a ratio of 4:5, how many would you have to pull out in order to guarantee a pair?” The instructions were read aloud to participants. However, a small-scale pilot indicated that the participants did not understand the question when posed in this way. A new set of instructions was designed for the main task which ran initially thus:

> In the bag is a brown sock and black socks and they are in a ratio of 4 to 5, so there's 4 brown socks for every 5 black socks. Is that clear? What I would like to know is the minimum number of socks that you would have to pull out at random in order to be sure that you would get a pair. You may pick up the bag and play with the bag however you see fit, you might find it helps you to pull out socks to help you figure out the answer and you have five minutes to answer the problem.

However, after five participants it became clear that additional information would need to be added which resulted in the instructions below:

> After I've explained the task you have 5 minutes to come up with an answer. If you come up with it before then, that's fine, just tell me. In this bag are brown socks and black socks in a ratio of 4 brown socks for every 5 black socks… so the ratio of brown socks to black socks is 4 to 5, is that clear? What is the minimum number of sock you think you need to take out of the bag at random to be sure you had a pair of either brown sock or back socks, so that is if you were to pull socks out at random from the bag, what is the minimum number of socks that you would need to be sure that you would have one pair, that's two socks, which match in colour whether that's one pair of brown socks or one pair of black socks.

While this slight change in instructions leads to an inelegancy in the procedure, the discussion of the qualitative results below will indicate that it made little difference to the actual quantitative results.

Participants were given 5 minutes to answer the problem and they were allowed to make multiple attempts. They were not prevented from speaking to me but nor were they requested to follow a think aloud protocol. I would only answer questions by repeating the relevant part of the instructions. Participants were allowed multiple attempts at the answer.

If they got a correct answer, I required them to explain why the answer was correct. If they did not manage to answer correctly, I explained the answer to them.

Materials and Measures

The participants' performance was measured in terms of success and the latency to a correct solution. If they answered the question correctly, they were asked to rate their feelings of insight along 5 dimensions. The same insight rating scale was used as in Chapter Five (taken from Webb et al., 2018). This required them to rate using a Likert-type scale ranging from 1-100 where 1 = low and 100 = high : (1) the confidence that the given response was correct ("very unsure" to "very sure"), (2) the strength of the aha experience ("very weak" to "very strong"), (3) the pleasantness of the insight experience ("very unpleasant" to "very pleasant"), (4) the surprising nature of the insight experience ("not surprising at all" to "very surprising"), and (5) the feeling of impasse before the insight experience ("no impasse at all" to "very stuck").

The researcher kept detailed notes while observing the participants; the participants were recorded throughout the study and the videos were later watched. The qualitative memos and coding were kept in Excel (see **Appendix H**).

Results

Data availability

The data reported and analysed here are available as **Appendix I** and an example of the coding is available as **Appendix H** in the online supplementary information. Performance was better in the high condition (Figure 6.2b) where 50% of participant solved the problem compared to the low condition where 38% of participants did so. This difference was not significant, $\chi^2 = 0.473$, $p = .492$. Cramer's V = .112. This difference remained when the solution rate for the first 2 minutes[87] was calculated (Figure 6.2a): 11% solved it in the low condition and 25% solved it in the high interactivity condition but this difference was also not significant, $\chi^2 = 1.216$, $p = .270$, Cramer's V = .179. Participants were slower in the high condition with an average latency of 116.20 seconds (*SD* =78.67s; although this was skewed by P41 who took 297 seconds) than in the low condition (*M* = 103.57, *SD* =48.32; Figure 6.3c). This difference was not significant, $t(15) = .376$, $p =.713$, Cohen's $d = .185$.

Participants who solved the problem were invited to rate the nature of their experience along five dimension; the mean ratings are plotted in Figure 6.2. Participants had less of a feeling of being stuck in the low condition (*M* = 45.00, *SD* = 40.56) than in the high condition (*M* = 69.40, *SD* = 31.50) and they felt more confident in their answer (*M* = 76.00, *SD* = 32.62; high *M* = 66.10, *SD* = 34.64) and found the feeling highly pleasant (*M* = 92.00, *SD* = 8.86; high *M* = 79.90, *SD* = 25.05). Those in the high condition were more surprised (*M* = 54.70, *SD* = 32.93; low *M* = 34.71, *SD* = 54.70) and both condition experienced similar levels of Aha! (low *M* = 66.00, *SD* = 32.84; high *M* = 66.50, *SD* = 24.95). None of the differences were significant, smallest $p = .182$

[87] This reflects the time given to participants in Chuderski et al (2020); see the discussion.

Figure 6.2

The Proportion of Correct Answers in (a) Two Minutes and (b) Five Minutes and (c) the Latency to Solution in Seconds (Error Bars Represent Standard Error of the Mean)

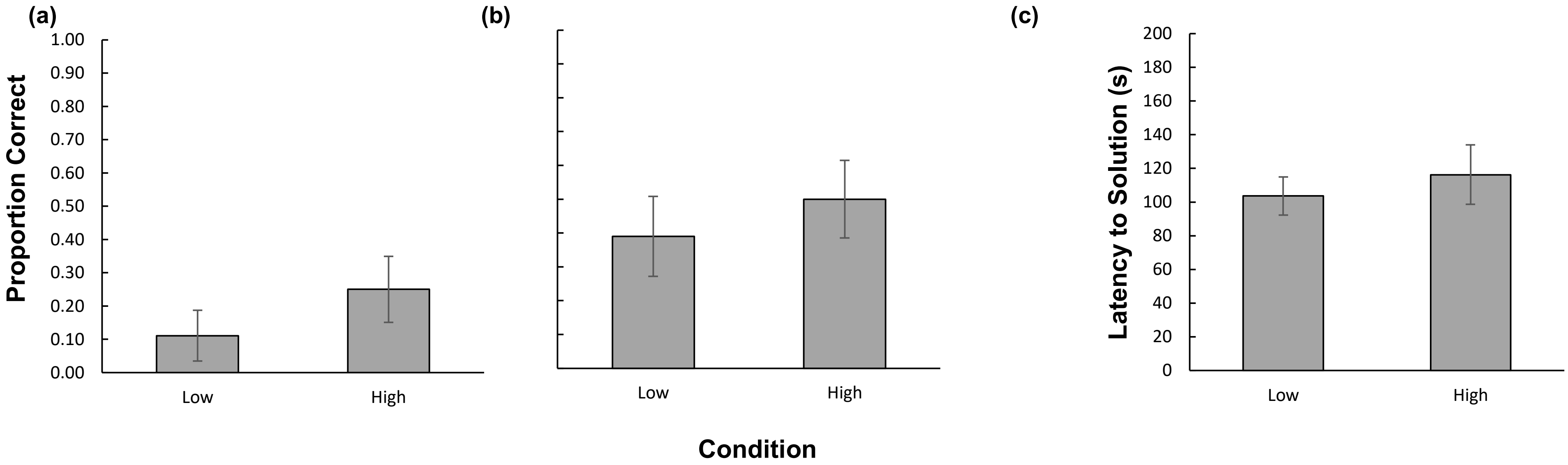
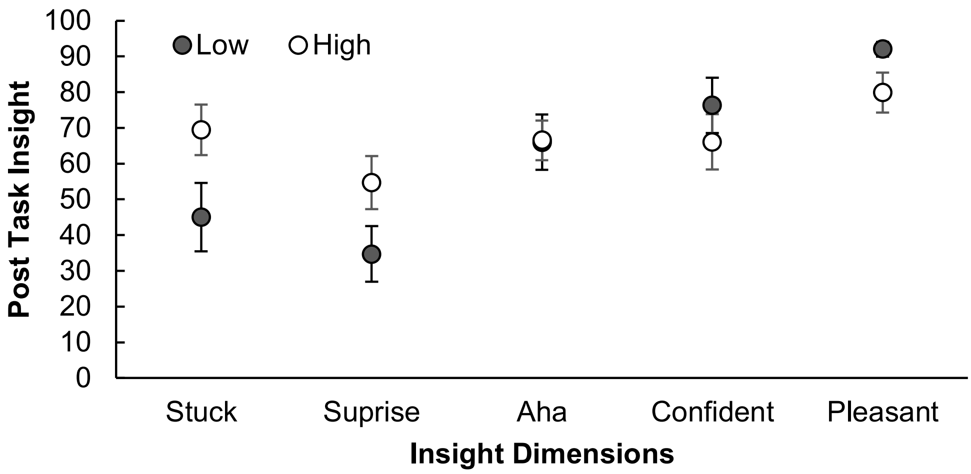


Figure 6.3

Average Insight across the Five Insight Dimensions Of "Stuck", "Surprise". "Aha", "Confident " and "Pleasant" (Error Bars Represent Standard Error of the Mean).



Qualitative Section

Method

As in Chapters Three and Four, the quantitative results outlined above were not in line with the hypotheses. This demonstrates the importance of measuring behaviour to support an explanation of why data can support or not the generated hypotheses. This third study was testing two hypotheses: a performance-based hypothesis and a behaviour-based one. Again, the initial quantitative hypothesis was not upheld – performance, while better in a high interactivity environment was not significantly so.

However, as the study unfolded, it became clear that the experimental manipulation failed and there were significant procedural and theoretical assumptions and errors which can be tracked through qualitative analysis. Some were failures in experimental procedures[88] which could be rectified in a second study (much as in Chapter Five) but as we will see, others undermined the underlying premise of the study and led to the decision not to repeat the study. The qualitative analysis reveals an inconsistent effect of experimental condition making it harder to draw clear conclusions.

Additionally, the method of kinenoetic analysis that was followed in Chapters Four and Five was not applicable as was originally planned. As will be demonstrated, the movement of objects did not allow the tracing of thoughts in the same way and resisted quantitivisation, while some strategies could be extracted from these movements, the main information came from spontaneous verbal utterances and while there were observations of physical movements, object-thought trajectories were frustrated and thinking did not occur in this way. I offer some reflections on the reasons for this below and the implications for a qualitative analysis of thinking.

As the researcher who conducted the study, it became clear to me that there were underlying patterns from which new knowledge about this problem situation could be deduced beyond that contained within the strictures of the hypothetico-deductive model. That is, that theoretically valid reasons for the practical and experimental failure of the experiment could be seen in the behaviour of participants throughout the experimental situation. The decision was taken to employ a modified version of Grounded Theory Method (Glaser & Strauss, 1967; GTM[89]) with the video data to generate novel theories about problem-solving in this task. Thus, the focus of the research shifted from whether

[88] In this respect filming the experimental situation is a useful tool to ensure adherence to a set experimental procedure. In this instance, there was only one researcher but in other cases there may be multiple researchers. Given we are still unclear of the contextual variables that may influence results (Leonelli, 2018) generating open ended data is necessary to understand the complexity.

[89] I follow Bryant (2002), Urquhart (2013) and others here in making a distinction between method and the theories which are generated by the method.

first order problem-solving environments support problem-solving assessed quantitatively to what happens when participants are placed in the environment.

Grounded Theory

GTM (Bryant & Charmaz, 2011; Charmaz, 2014; Glaser & Strauss, 1967) seemed a particularly appropriate framework for the analysis of the data. It has the same pragmatist roots as kinenoetic analysis (see Chapter Three) and aims to start from an assessment of the data rather than generating theories. Urguhart et al. (2009) identify four key characteristics of GTM: First, it has as its main purpose theory building. Second, it requires a theoretical naivety from the researcher avoid preconceived ideas. Third, data are collected iteratively and compared with all existing concepts as the theory grows. Fourth, slices of data are collected through theoretical sampling where more data are sampled to strengthen the growing theory.

GTM requires a theoretical naïveté so the researcher is not tempted to undertake a top-down analysis of the data. This is particularly salient in this case because the qualitative analysis was undertaken specifically because the existing theories did not explain the performance in this task indeed, they predicted the opposite. Therefore, the inductive and generative approach associated with GTM could help to make sense of this theoretical vacuum. Indeed, while a traditional GTM analysis cautions against a familiarity with the existing literature allowing the researcher to approach the data in a naïve fashion, the underlying lack of theoretical explanation for the phenomenon observed meant that this uninformed epistemic state developed naturally. In this sense the work described here fits both aspects of the surprise value of research outlined by Gaskell and Bauer (2000, p. 347) “either with regard to some common sense view, or with regard to some theoretical expectation." There was no literature of which I am aware which addresses the nature of interaction with objects in an experimental situation in this way to provide the theoretical scaffolding. All starts from the hypothesis that interactivity is a beneficial scaffold[90].

The iterative nature of GTM also suited the process of research in this instance. The initial observations were made in vivo (noted by me as a researcher as part of normal research practice) and could be followed up with theoretically motivated sampling of the video data in a recursive manner as underlying concepts arose. While this was an unintentional epiphenomenon of the experimental set-up, it also meant that the process of data analysis reflected that of GTM. Additionally, the underlying positivist leanings and claim for data objectivity which is held to by traditional GTM was important because the notion of the participants' understanding was not part of the research question of this study. Therefore, while I acknowledge my own theoretical position and the influence this may have on my sense making there is no underlying double hermeneutic: The data here are not doubly constructed and therefore follow a more traditional GTM approach (Urquhart et al., 2009)

The study diverged in a surface way from the notion of theoretical sampling because the data collection only happened at one time point rather than involving a return to the participants to collect data which might expand on the nascent theories (theoretical sampling). However, in practice the recursive nature of the data analysis from initial in vivo observations, to behavioral coding to conceptual construction meant that a form of

[90] This is not to say that the interpretation was completely theoretically naïve. As argued by Kelle (2011) no data can be produced in a theoretical vacuum and it was impossible for me to shed my allegiances to a systemic approach to cognition. Indeed, this is why the qualitative phase was carried out in the first place – a different theoretical allegiance would have meant I accepted the quantitative results as they stood. By theoretical naïveté here, I wish to stress more that the particular phenomenon that I wished to examine has no underlying specific theoretical justification, therefore I was forced to approach the data inductively.

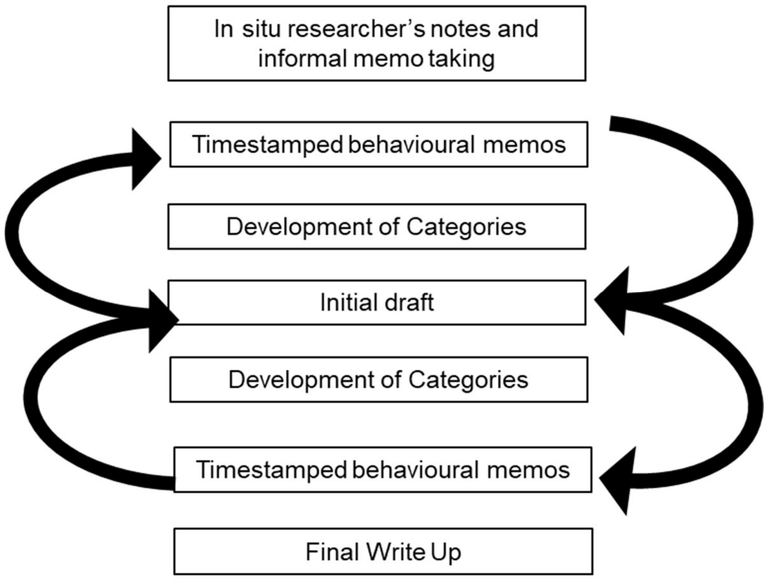
theoretical sampling emerged. That is, those participants who were most relevant in that situation had their video data reanalysed in the light of the emerging theories. This occurred until saturation was achieved. Video data even outside of GTM often elicits this type of data analysis because the very rich level of the data requires the researcher to selectively code in the open coding phase and revisit the data later to resample in light of conceptual developments (Heath et al., 2010).

Analytical Process

The initial observations generated in vivo generated were followed up by close watching of the video data with time stamped qualitative memos. Conversation was transcribed and also time stamped. These were then grouped in conceptual categories before the videos were watched a second time to substantiate these categories (this is the modified theoretical sampling which while did not generate new data did return to the existing data with modified questions). This iterative process continued through initial drafts of this chapter. This process is illustrated in Figure 6.4 and an example can be found in **Appendix H**. In tandem with the data analysis carried out through watching the videos, the results were discussed in weekly meetings with my supervisor and conceptual categories were refined.

Figure 6.4

The Iterative Analytical Process. The Process Moved from InVivo Notes and Informal Memos to Time Stamped Behavioural Memos Which Were Refined as Theories Were Generated and Substantiated through Recursive Theoretical Sampling



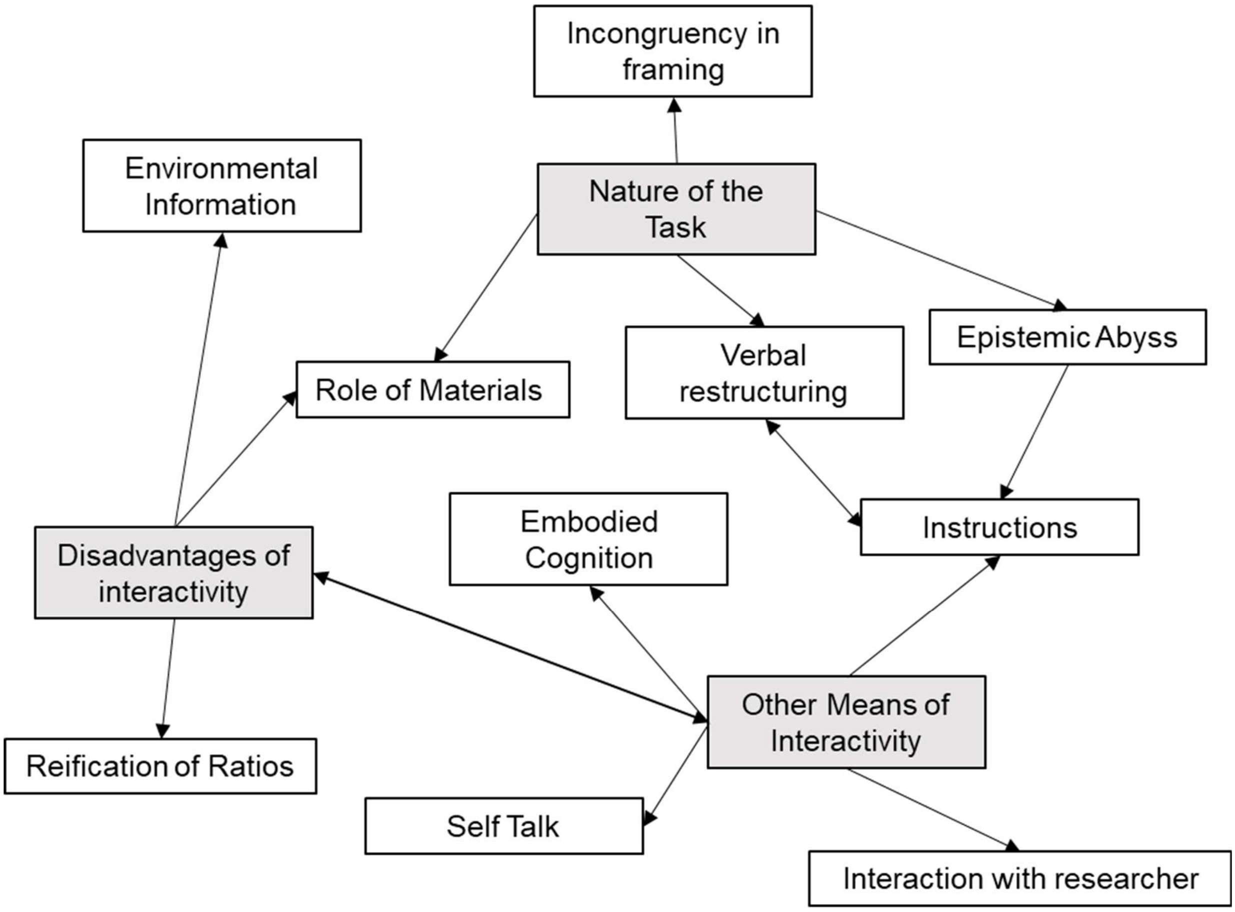
Results

The benefit of video data is the rich interplay of movement, words and interactions than can be mapped. This is also a disadvantage. I therefore take quite a narrow focus on the qualitative data and use it to explore explanations for the inconclusive quantitative data.

The initial observational categories generated in vivo are highlighted in a darker colour in Figure 6.5: They were the failure of interactivity to scaffold problem-solving in this task, the confusing nature of the task which meant the participants suffered almost paralysing at times epistemic abyss which I defined as an inability to find a starting place to solve the problem, that there were other forms of interaction. These initial observations were expanded through repeated viewing and iterative sampling to produce the conceptual map in Figure 6.5.

Figure 6.5

Conceptual Map Based on Initial In Vivo Categories and Repeated Video Sampling. The Dark Boxes Represent the Initial Observations Generated In Vivo, The Additional Categories Those Which Arose from First Review of the Video.



The arrows indicate the extent that the conceptual codes overlap with each other and form bidirectional relationships. It became necessary to reduce the codes to substantive theories and those that answered the initial research question: why there was no significant different between the conditions. The most important concept to emerge from the analysis was the observation that interactivity did not scaffold problem-solving in this study—contrary to the original hypothesis—and a closer analysis pinpoints why this might be so:

An excess of environmental information and the encouragement to interact with it reified unhelpful representations. Alongside this I note than another form of interactivity arose: I observe a form of linguistic interactivity which was focused on the instructions and the relationship with the researcher. The iterative process of discovery emerged more often in this linguistic space rather than in the space between object and person. This interactivity was dispersed across experimental conditions and was in part the reason for the inconclusive results and the failed manipulation. Therefore, the initial cluster of categories displayed in Figure 6.5 was reduced to these two broader theories in line with the theory development stage of GTM.

Information in the Environment

For Kaplan and Simon (1990, p. 390) the added information from the real world supports problem-solving (in this case with the mutilated chessboard problem[91]) and reveals new problem affordances much like the theoretical augmentation posited by interactivity:

> Using information from the real world, the simulation is able to shift from an initial representation of "square," to a representation of "black square" or "white square." A similar production allows the simulation to elaborate old propositions using the new concepts of colored squares. Thus, the proposition "A domino covers a square and a square" becomes "A domino covers a black square and a white square."

This section of the results suggests that this information can act as a barrier to problem-solving by offering and reifying multiple unhelpful pathways.

91" Suppose a standard 8×8 chessboard has two diagonally opposite corners removed, leaving 62 squares. Is it possible to place 31 dominoes of size 2×1 so as to cover all of these squares?" The puzzle is impossible to complete. A domino placed on the chessboard will always cover one white square and one black square. Therefore, a collection of dominoes placed on the board will cover an equal numbers of squares of each color. If the two white corners are removed from the board then 30 white squares and 32 black squares remain to be covered by dominoes, so this is impossible. If the two black corners are removed instead, then 32 white squares and 30 black squares remain, so it is again impossible.

The socks problem is a hard problem because participants get stuck on the unhelpful ratio. The structure of the question invites this initial representation with its focus on proportions and its invitation to use a probability formula (Weisberg, 1995). As suggested in the original behavioural hypothesis from F. Vallée-Tourangeau and March (2019), ignoring the ratio should lead to participants solving the problem more often and faster because they are not tempted to use inappropriate mathematical formulas. This is not unreasonable. However, the specific hypothesis under exploration, that a high interactivity environment will increase the number of people who disregard the misleading and uninformative ratio is not sustained by an analysis of what people actually do in this condition, the opposite seems to be the case, people use high interactivity to reinforce the unhelpful ratios. As we have seen in Chapters Four and Five, a high interactivity environment can scaffold unhelpful representations as well as helpful ones. Indeed, the idea that the environment would only scaffold helpful ones harks back to the omniscient problem solver who would knowingly select the correct representation. There are two ways that a problem solver could approach solving a problem using interactivity: The first involves the sort of exaptative actions outlined in the intuition pump described above, the second involves using the world to scaffold existing representations. In the case of problems which are designed to elicit an unhelpful representation, it is plausible that such a representation when it is enacted would be harder rather than easier to disregard. Indeed, the reification of a representation and its solidification is one of the benefits of interactivity.

This reification was clearly in evidence in the study reported here: Interactivity allowed participants to represent in a solid form the ratio. Many of them pulled 5 socks from the bag and then 4 black socks and used these to structure their thoughts so the ratio became more, not less, concrete and therefore harder to disregard. Just as not every representation is a helpful one, neither is every strategy helpful. Some of these strategies are more unhelpful than others. Take for example P5 who took socks from the bag and held them above her head on several occasions during the task (Figure 6.6)

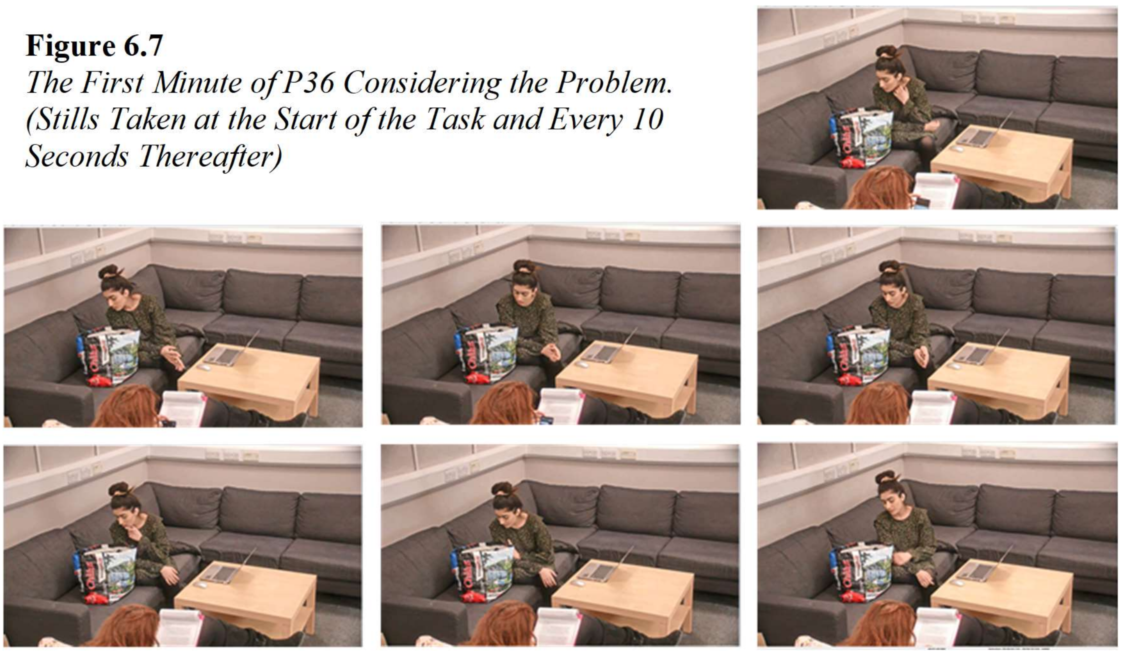
Figure 6.6

P5 Pursuing the Unhelpful Strategy of Taking Socks Out of the Bag and Holding Them in the Air (Still From 00.59).

This was unlikely to widen her epistemic field. However, while some strategies are clearly not appropriate, many other were an objectively sensible strategy from the perspective of an ignorant problem solver. It is not unreasonable to assume that the information in the problem is important, it would be incongruent for a problem to specify these ratios unless they were important especially perhaps when presented in conversation. Therefore, the strategy of reifying the ratios is both sensible and yet paradoxically makes the problem harder to solve.

However, more than allowing the reification of the information in the problem, the presentation of actual sock expanded the amount of information and unproductive pathways. Problem-solving in this less controlled environment did not scaffold thought, rather it created an additional distraction. This was perhaps reinforced by the instructions which directly called the participants' attention to the bag: "Take this bag, in it you will find black socks and brown socks". Indeed, the concrete anchoring of the problem led people to look at the distribution of socks across the actual bag that they had in front of them rather than considering an abstracted entity. Kaplan and Simon (1990, p. 377) suggest that clues in the environment serve to "constrain search" but they also open up possibilities which are not present in a second order environment. This was true even of those that were in the low interactivity condition who tended to return their gaze to the bag when thinking. Take for example, P36. Figure 6.7 demonstrates the extent to which she is looking at and into the bag even while not interacting with the socks in it. Throughout the course of the experiment, she looks at the bag and the accompanying finger gestures and mouthing of number indicate that she is trying to count the socks.

Figure 6.7
The First Minute of P36 Considering the Problem. (Stills Taken at the Start of the Task and Every 10 Seconds Thereafter)



So, rather than supporting problem-solving, the actual physical arrangements of socks added additional complexity. P26 (low interactivity) said “Well looking at this bag, all the black socks are at the top so it would take a while to get a brown pair. Oh, I’m so confused” (01:36). Within a traditional problem-solving environment, the participants know that they are solving a riddle and they are in an artificial situation, in this experiment the mundane situation of finding a pair of socks was collapsed into a school type problem. There simply is no “real world” alternative to the hypothesised riddle. If you want a pair of socks you get them out of the bag. Anything else is an artificial constraint. Making sense of the parameters of that constraint was made harder by the experimental situation

Additionally, while it is certainly true that the participants were offered much more information when they were presented with the bag of socks in both the low and the high interactivity environment and, more than that, in the high interactivity environment they were further offered the opportunity to interact with that information, this did not benefit them in any way. Take for example the number of socks that are in the bag. This information is not salient in a second order problem-solving environment but when the participants are presented with bag full of socks and a problem masquerading as a mathematical problem, as we have seen the total number of socks takes on more importance.

Investigating this unproductive cul de sac was easier for those in the high interactivity condition and therefore harder to disregard. For example, P37. After asking for clarification of the problems at 33 seconds into the task he asks: “How many were there in total?” (00:33). After 58 seconds have elapsed, he decided to empty the bag and count the socks. It is hard to overstate what a poor decision this and it is perhaps best illustrated

Figure 6.8

P37 Tidying Socks (Stills Taken at 60 Seconds and Every 10 Seconds Thereafter)

by Figure 6.8 where a screenshot was taken every 10 seconds over the following 220 seconds. Perhaps like P25 who states “Ah, that’s what I have to do, I have to know how many are in here” (02:15) before tipping the bag out, the participant thinks he has solved the “trick”. Whatever the reason, the messiness of tipping out all the socks and counting them and the high time cost meant that he made no progress. This wrong route would not have been possible in a second order problem-solving environment. The information simply does not exist to consult. This has important practical implications for problem-solving in the flux of people and things that make up everyday life outside the laboratory, the skill or luck may lie as much in deciding which information is important as it does in using that information to inform and structure problem solving.[92]

Porous Cognitive Boundaries

Over the course of the analysis it also became clear that there were versions of cognitive interactivity which were not related to the moving of the objects and yet which scaffolded the pathway to solution. The instructions played a critical role in this scaffolding. This was an unintended artifact of the experimental situation and a serendipitously inelegant one. The instructions were given verbally, and the participants were allowed to ask for clarification. The clarification took the form of the experimenter repeating the relevant set of instructions. Thus, what is often unspoken (checking instructions) was given a more traceable form and this allowed us to track this experimental artifact.

What this underscored was the importance of the instructions in the problem

[92] This is perhaps an example of how scaling up which is discussed at greater length in Chapter 7 is not the problem. Rather cognitive resources in a real world problem-solving environment are as much devoted to the identification of the problem as its solution. Insight problem-solving is really an example of this, find the right problem and the solution appears.

solution. Initially, there was a change made to the instructions as outlined in the method section in order to ensure all salient information was clearly presented but this did nothing to change the participants' need to check. It soon became clear that the participants did not understand the problem as presented and would recode it unhelpfully. The instructions were initially changed but this did not stop the participants asking repeatedly for clarification. This clarification revealed two things: First, a consistent epistemic state of the participants should not be assumed and second, that it was possible to use this to structure understanding. Indeed, the nature of the task is to restructure the instructions to gain understanding rather than restricting anything in the visuo spatial space, therefore with hindsight[93] it is not unreasonable that the interactivity coalesce on this linguistic plane.

The verbal nature of the instructions also increased the interaction with the researcher commented on above. While this is an admitted inelegance of the research design, it revealed the need for the participant to seek help from the outside world. Over the course of these interactions, the answer often revealed itself through a gradual and recursive process. Take for example, P41 (Figure 6.9) the last minute or so of her problem-solving involves a recursive, back and forth between extracting information from the socks that she has in her hand, the instructions and the researcher through making guesses until she suggests the correct answer. Note how often she looks at the researcher in the episodes selected and elicits her feedback. As argued in F.Vallée-Tourangeau et al (2020), pinpointing the moment of "insight" without documenting all the steps that have led up to it gives a distorted view. In this case, playing with socks coupled with the information from the researcher and the redrawing of the epistemic map through the wrong guesses,

[93] Hindsight always implies a projected thought which is transformed in action to reveal a different outcome. Hindsight only exists after action.

scaffolded the way to a solution.

Figure 6.9

P41 Solving the Problem

04:53	Pulls out two more - this time both black	
05:07	"This is 8....."	
05:08	Looks at researcher	
05:14	Pulls through socks in hands	

05:21	"So I have to pull out a pair?"	
05:23	Research repeats instructions	
05:35	"So I could pull out four socks and then two would be a pair"	
05:42	Researcher repeats request for the *minimum number*	

05:45	Plays with socks in her hand	
05:50	"Three"	
05:57	Correct explanation	

This was even clearer in the low interactivity condition where the participants sought out scaffolding more obviously. Take for example, P26 in Figure 6.10. This was a low interactivity participant who uncovered the answer through series of guesses, much like the participants in Chapters Four and Five would incrementally uncover the answer through shaping the array. As she states clearly when asked if she knows why: "When I got to the end I did...when I said two and then I realised...." (02:51). The act of saying the word alongside the feedback from the environment scaffolded her understanding. Outsight occurred when the thought was in the world even if that outsight did not take material form, it was still generated in action, this time the action of guessing. To place the results from this participant in a low interactivity environment would be misleading even if she did not carry out any actions over objects.

Figure 6.10

P26 Solving The Problem Through Incremental Guessing

01:18	"Oh what and if I get it right you'll tell me?"	
01:21	"Hang on, I don't know how many socks are in the bag?"	
01:24	Looks at researcher expectantly.	
01:27	"So I don't get how the ratio would matter if I don't know how many socks were in the bag....oh. it's just chance isn't it?"	

01:36	"Well, looking at the bag actually all the black ones are on top so it might be a while to get a pair of brown ones.."	
01:45	"I'm so confused by the question" Looks at researcher again	
01:47	Researcher repeats instructions	
01:50	Starts to pull socks; Researcher stops her	
02:03	Researcher ends instructions	
02:11	"I'd have to pick 10 socks out"	

02:17	"I'd have to pick 8 socks outs"	
02:23	"I'd have to pick 20 socks out"	
02:32	"I'd have to pick 18 socks out"	
02:41	"I'd have to pick 4 socks out". Said faster and more determined	

02:44	"2 socks"	
02:47	"3 socks"	
02:49	Researcher confirms	
02:50	Starts to laugh	
02:51	"When I got to the end I did...when I said 2 and then I realised...."	

Transformation from Second to First Order

The original problem as presented by the prior literature cited in the introduction reads: ‘You have two kinds of socks in your drawer, mixed in a ratio of 4:5. How many socks will you have to take out to be sure of having a pair of the same kind?’ Pilot trials revealed that this question was not suitable to a study where participants were given actual socks to take out because the answer was simple. Two. The problem is no longer a hypothetical one but one where there are socks and drawing a pair is simple. Even with the clarification in the instructions, participants struggled with the moved from second order to first order. As P15 says "But my question is, can I look at the socks because if I can see the socks then I can pull it immediately" (01:44). In a problem presented simply as a thought experiment, the socks cannot be seen and tangibly grasped. In traditional presentations, the problem is framed as a riddle, in the way the problem was presented here the problem becomes a practical task and as such an easy one. Two was a common guess across the participants and many simply became confused about what constituted "a pair"[94]. Thus, they turned to understanding the instructions and the scaffolding observed above occurred.

Discussion

The quantitative results were inconclusive. It may be that if the study had reached the anticipated number of 60 participants, they would show a significant augmentative effect of interactivity but it seems unlikely from the direction of the results And indeed the cognitive messiness of the procedure would mean that it would be hard to trust that the results reflected the experimental manipulation[95]. The problems I have outlined above by the standards of traditional experimental psychology would cast the experiment as a failure. As argued by Gozli (2017), the experiment in psychology takes the form of rule governed behaviour. He uses the example of a chess player pouring hot coffee onto an opponent's player in a chess game as an example of how the game is not won, the player is not playing by the rules. In the experiment reported here, the participants did not play by

[94] The confusion around what constituted a pair appeared to represent a desire from these participants to somehow make the question more difficult.

[95] An important consideration. Not only does tracking behaviour allow us to understand the mechanisms behind the quantitative results, it also allows us to avoid "false positives" by ascribing quantitative results to the experimental manipulation when they may have other causes.

the rules of the experiment: Namely to use only brain based operations in the low interactivity, sequestered condition and to make full use of the objects in the high interactivity condition. Thus, the experimental manipulation and the experiment failed.

Furthermore, it is unclear how useful the quantitative results would be without an understanding of the mechanisms which underlie interactivity. Certainly, with the evidence that we have discussed above that seems unlikely. The qualitative analysis undertaken here, however, starts to offer an understanding of the detrimental role that interactivity can play as alongside the augmentative role traditionally theorised. In hindsight such an effect is theoretically more plausible given an open cognitive system and an enactivist, non-computational approach.

Methodological solipsism relies on two things, that a cognitive process can be meaningfully studied in isolation and that it is possible to isolate the processes. The contradiction at the heart of experimental research which contrasts a low and a high interactivity condition is that this ignores the other aspects of the system. The data presented here suggest that a system forms around any problem solver and the salient aspects are not necessarily those which are being tested once we approach cognition from a systemic perspective. Here when interactivity with the objects proved to be unhelpful that interactivity seeped out into the wider system to encompass the interaction with the researcher via the instructions and the multiple opportunities to offer solutions.

Chuderski et al. (2020)

During the running of the study reported here, Chuderksi and colleagues (2020) published a large-scale replication of many of the experiments in interactivity and insight including the sock problem. Their aim was to test what they called the “interactivity hypothesis" which broadly follows the assertion outlined above that interactivity will lead to increased problem-solving success. Their overall conclusion was that interactivity has no effect on problem-solving (although see a later paper from the group which claims that interactive formats to "increase the chance for genuine insight" [Chuderski, Jastrzębski, Kroczek, et al., 2020, p. 4]). They measured the effect of interactivity in terms of differences in solution rates. In the case of the sock problem, participants were asked to generate the solution either using only pen and paper or were given a cardboard box with socks. A Bayesian analysis suggests that the evidence points to an equality between the groups. There was no further detail of process or level of engagement with the task in these conditions.

The quantitative results of the study in this chapter support these results. There was no significant effect of interactivity across the conditions. However, the qualitative results do not support the conclusion that interactivity has no effect. Rather, it has an inconsistent and idiosyncratic effect which would be hard to demonstrate using the evidence generated by Chuderski et al (2020). A later trial comparing pen and paper reported in this paper and a completely non interactive showed a similar lack of difference and also showed that only 12 of the 93 used the pen and paper. This window into process support the idea that interactivity however limited is detrimental in this task. Of the 12 problem solvers that used pen and pencil, only 2 got the correct answer. Given that 26.4% of the participants in the low interactivity condition and 24.7% of those in the high who did use pen and paper did not solve it, the success rate of 16.6% seems low although, just like in Chapter Five it is problematic to place participants in these post hoc groups.

The only way of determining the effect of interactivity on solution rates is to include a manipulation check. It is as puzzling for Chuderski and colleagues to compare solutions across two groups of participants where only 12% of participants in the experimental condition requiring interactivity actually engaged with that format. This is the same argument that was made in Chapter Four with regards to the work on word production reported in Maglio et al., (1999) which revealed that many people did not use their hands. It is worth noting in this respect that Chuderski et al., (2020) invited their participants to solve 9 separate insight problems. While for the harder six, participants were asked to create the answers from the materials they were given thus forcing a level of interactivity no matter how slight, for the easier three they were invited to offer a verbal response. This suggests that the researchers did not expect the participants to interact with the materials in these "easier" questions.

Experimental Research in Interactivity

Complex systems interact in complex ways and the results are emergent and contingent. This thesis so far has focused on the movements of objects in the world and this is still the focus of the research programme that I will outline in Chapter Eight where I will make the argument for the movement in the world of action-thoughts. However, soft assembled cognitive ecosystems which form and reform make the best use of any resources around them. Careful thought needs to be paid to the experimental situation so that the co-construction of the problem solution can be easily witnessed and tracked.

For example, in F. Vallée-Tourangeau et al (2020) participants were invited to solve the triangle of coins problem using a touch screen computer. Half the participants were prevented from moving the digital representations on the screen and were asked to mentally simulate the rotation and half are allowed to move the tokens. The first experiment in this paper allowed participants to guess as many times as they liked. In the low interactivity condition, participants made more guesses and used the information from those guesses to help structure their thoughts, there was not a lack of interactivity but rather the type of interactivity shifted from moving the tokens to retrieving information from the incorrect announcements. In other words, rather than low interactivity, the type of interactivity changes. This echoes the relationship between the instructions, guesses and researcher interaction presented above. It is simply unknown to what extend this level of interactivity affects other experimental situations. Indeed, with much research moving online we shall potentially lose the opportunity to ever measure this.

Just as in thought experiments, pre specification of underlying causal mechanisms is difficult because an adherence to a form of wide cognition means that such pre-specification is necessarily an under specification. Pickering (1995, p. 24) suggests that discovery (which in this instance we can read as discovery of the answer) passes through a mangle of practice which makes causal and predictive explanations impossible:

> We just have to find out, in practice, by passing through the mangle, how the next capture of material agency is to be made and what it will look like. Captures and their properties in this sense just happen. This is my basic sense of emergence, as sense of brute chance, happening in time – and it is offensive to some deeply ingrained patterns of thought. The latter look for explanations—and the closer to the causal, mechanical explanations of classical physics the better—while it seems to be that in the analysis of real-time practice, in certain respects at least none can be given.

Toomela (2020) argues that definitions in psychology are unstable because they are not anchored to a material reality and while I have some sympathy with this view, actions in and on the world provide a material reality for the anchoring of psychological concepts and definitions. I would therefore argue that human behaviour is not an ideal type that can be predicted through the use of complex cognitive modelling or (however well crafted) thought experiments but is rather an ongoing and idiosyncratic process. However, as

Pickering also goes on to argue, patterns emerge from the descriptions we generate from practice.

This point is also made by F. Vallée-Tourangeau and March (2019) who use the data from the 17 animals experiment to emphasise the "singular and contingent" (p. 5) problem-solving trajectory which also took participants down unhelpful and ultimately blocked routes such as building high nests or, in one case, using the clips provided to attach the animals to the models. Both these choices obscured the true route to solution. However, unless we are dealing with the sort of omniscient problem solver outlined in the introduction, it is impossible to predict in advance which solution routes will be useful especially if that problem solution turns on the sorts of unplanned and exaptative gestures chronicled in Steffensen et al., (2016). Recognising this does not mean that research in interactivity is has somehow failed but rather that it should turn to a consideration of how the correct route is chosen when it is not immediately apparent. It suggests that in insight tasks where there is a double ignorance—ignorance of solution and ignorance of path to solution—that both the finding of the solution and the path taken become important.

Interactivity posits that a dynamic relationship with the external world is not only augmentative to cognition processes but necessary. However, first wave experimental research in interactivity does so by setting up a condition in which it these things are not necessary. In this way it adheres to the fallacy of the unreal world, that is that there is a way to compare non systemic and systemic cognition and that non systemic cognition can be isolated in an experimental psychologist's lab. Rather, I propose that all cognition is best approached as system but the nature of the systems that form are different.

The Instrumentalization of the Problem Space.

KA relies on a well instrumentalised problem space. The main times it has been used is in Chapters 4 and 5 of this thesis and in F. Vallée-Tourangeau et al (2020) and in each time the tokens were not only abstractions (tokens standing in for something else – letters or coins) but they were also labelled and served specific functions in relation to the problem. The socks in this problem were interchangeable and did not function as external representations. As we have seen above this encouraged a direct relationship with the experimental material but it also frustrated attempts to systematically track the cognitive path through the movement of object-thoughts. In this instance, the solution does not rely on the fortuitous arrangement of tokens but rather a reframing of the problem. Objects in this instance did not scaffold understanding in the same way because crucially even if the correct number of socks were drawn this was not immediately apparent.

The Nature of First and Second Order Interactivity

This also indicates that there is a certain naiveté in assuming a low cost of translation from second order to first order problem-solving. The idea that problems can be moved from an abstracted problem space to a materially rich one assume the most important aspect is the movability of the objects. In the studies presented in Chapters Four and Five, the letters had a role as external representations offered the three benefits of offloading, scaffolding and suggesting. In other interactivity research, the movable objects are still abstracted representations of the objects in the puzzle, from shapes on a tablet standing in for coins (F. Vallée-Tourangeau et al., 2020), to cards standing in for people (G. Vallée-Tourangeau et al., 2015) or links standing in for a whole necklace (Fioratou & Cowley, 2009). Thus, while the objects became tangible, the problem space still did not. It is as if the17 animals experiment in Steffensen et al (2016) were translated to a field with real animals and real pens. Just as the material affordances matter (as suggested in Chapters Three and Four) so too do the overall task parameters. [96]

Thus, there is a further naiveté in the conception of first order problem-solving in Vallée-Tourangeau and March (2019) mirrored in my initial experimental set up, that is that the translation from first to second order can occur across modalities. The sock problem is not of the same type as either the word production or anagram tasks. Unlike the 17 animals problem, the answer cannot unfold unproblematically in front of the problem solver until she can already recognise it. Drawing a brown, a black and then a brown sock did not trigger the understanding that uncovering a word stem or overlapping pens would do. Understanding the problem required making sense of the question which further required verbal and linguistic support rather than the support of objects. Viewed in this frame, it makes sense that participants abandoned unproductive actions over objects and rather sought scaffolding from the instructions and the researcher.

A systematic approach to cognition requires that every thoughtful (an arguably the non-thoughtful ones) encounter should be analysed as part of a system however what constitutes part of the system is rarely specified. F. Vallée-Tourangeau and March (p.3) suggest that a cognitive ecosystem is configured by: "the reasoner, the physical reality of

[96] There are various ways, the task parameters could have been altered to scaffold success in this task. It would be interesting for example to assess whether the use of sock tokens would overcome this uncomfortable incongruence between abstract problem and concrete objects or whether the instructions could be altered to restrict the range of possibilities. The system could also be constrained to restrict actions in the high interactivity condition such that only one sock at a time can be pulled out of the bag, and participants prompted to narrate the results of their action after each draw. However, this would still adhere to the view that the success or not of this experiment resides in whether it can be altered so as to improve performance. The argument developed in this chapter is that problem solving researchers interested in interactivity should not be so focused on the role of interactivity to engineer the circumstances to achieve a normative standard of performance, in this instance announcing the correct solution to this simple riddle, but rather that their research practice be more attuned to all the properties of the environment and the chain of actions they afford and from which thinking emerges.

the problem, and the action possibilities offered by the external environment". I argue that the data here suggest "the action possibilities offered by the external environment" are broader than whether the problem is represented in movable tokens or not. Rather participants actively recruit beyond the resources of interest in the experimental conditions.

Current experimental research in interactivity is interested in the introduction of movable objects into the cognitive ecosystem and therefore sets up two different experimental conditions: One with movable objects and one without. There are several problems with this approach to generate evidence for anything beyond a theory of the embedded mind (see Chapters Three and Five for a more in-depth discussion of this), yet the evidence is often used to support theories of more cognitive intermingling. First, the data here suggest that the behaviour of the two groups in this experiment is not that different. If this is the case, the ascribing causal mechanisms to a movable problem representation is simply not possible. Second, it upholds a residual dualism which leaves it both vulnerable to attack and at times posits contradictory theoretical positions.

In Chapter One I suggested that the binary between low and high interactivity sustained the fallacy that there was an "unreal world", that is one where the putative cognitive processes would be able to be isolated. As I outlined above, research in interactivity aims to establish that an embedded reasoner would reason better that one who is relying solely on mental processes. The video evidence here suggest that the parameters of the ecosystem expand when necessary and do not just include the movable objects, breaking the clear binary between low and high interactivity. Indeed, in this case while the objects yielded little useful information, the participants sought other forms of scaffolding.

This requires us to rethink how we do experiments on interactivity. Once the subject of analysis is behaviour rather than performance measures and once we establish that adherence to experimental conditions differs across participants, the nature of the research necessarily shifts. We are no longer straining to establish an augmentative effect to "justify" the research, rather we are unveiling thought with all its flaws and inefficiencies.

Conclusion

The data presented here are not the clean data that are typically seen in psychological reports of experimental situations. The cleanliness of such data undergird much of the abstracted cognitive models where performance is unambiguous. I suggest that the participants in cognitive experiments are ideal types much like we posit in thought experiments working in abstracted problem spaces. Even if the problems space is crafted so that a more direct attention is given to the interaction with objects or artefacts then other forms of interactivity are ignored. However, it is often impossible to tell in advance where the patterns of interest lie as I have established in this chapter and the two proceeding.

Over the course of the study reported here, a more ethnographic approach to the whole experimental situation highlights the messiness of in laboratory research. Some aspects are particular to the situation of an exploratory, quasi pilot study and others are broader reflections on the nature of research in psychology more generally. As we have seen above, the transformation from a second order to a first order problem space does not necessarily occur without cost and the nature of that cost deserves to be explored. It suggests a resistance to a reductionist approach. In Ross et al.,(2019), we argued that the low interactivity condition was controlled whereas the high interactivity condition was less so which limited the usefulness of firm conclusions. The data presented here suggest that both conditions are porous.

Traditional research in cognitive psychology aims to isolate the problem solver from the environment to better isolate the underlying cognitive processes (methodological sequestering; F. Vallée-Tourangeau & Vallée-Tourangeau, 2014, see also Chapters One and Three). However, it is often hard for it to not creep back in even in purportedly fully sequestered studies. This suggests that a true analysis of cognition should step outside of the boundaries of experimental research to assess underlying possible parameters before going back into the laboratory. Indeed, the second part of F. Vallée-Tourangeau and March (2010) reverses this and takes the insights from the laboratory and applies them to the work of a ceramicist gaining inspiration. The next, and final, empirical chapter of this thesis will move to the ceramicist's atelier to track the blooming of a thought.

Chapter Seven : From Kinenoesis to Kinemorphosis: Creativity in Action

Laboratory problem-solving tasks work best with well structured, knowledge lean questions. These are the sorts of problems are presented in the first three studies. Chapter Four looks at a word production task which is more likely to be considered to be solved through analytical methods although the finely grained case study demonstrated that there were moments of insight during the process, Chapter Five presents a hybrid problem (anagrams) and Chapter Six examines performance on a classic 'insight problem'. Each are well-established problem-solving tasks.

They bear little relation to the problem-solving we see in a more mundane environments because (as outlined in Chapter Three) they are predicated on the idea that problem-solving is modular and can be reduced to non-environmentally dependent processes. To better focus on those processes, experimental psychologists reduce the moment of interest to an internal, cognitive change. It is the argument of this thesis that such a sequestered take tells us little about problem-solving as a dynamic process situated in a varied landscape. And daily life tends to take place against the backdrop of such a varied landscape.

The data presented in this thesis to this point have demonstrated that a focus on process rather than outcomes reveals idiosyncratic and contingent pathways and collapses the distinction between creative and analytical problem-solving. Rather the process is not determined either by the problem or the affective response. While Chapters Four and Five explore classic laboratory studies which migrate easily to a more interactive format and are well instrumentalised to allow a form of kinenoetic analysis (KA), Chapter Six revealed the limits of transferring studies in this way and in so doing questioned whether such a move would be automatically useful: The transformation of an abstract problem into a first order space changed the form of the problem so much that it no longer retained some of its key qualities. The "school-like" nature of the problem introduced a layer of complexity that was not easily resolved.

The study presented in this chapter is an exploration of how creative problem-solving could be approached in a different way using a focused ethnographic and qualitative method while still drawing on the microgenetic basis of KA. In this chapter the focus shifts from how knowledge emerges from action to how form emerges in action; from kineneosis to kinemorphosis. It reports the process of making a clay flower and examines the problems which are encountered and solved along multiple timescales. It suggests that creativity arises from the act of problem-solving whether those problems are intentionally generated or naturally occurring, rather than a linear model which posits creative thought as anterior to solving the problem. New ideas, new objects are created through action rather than prior to it.

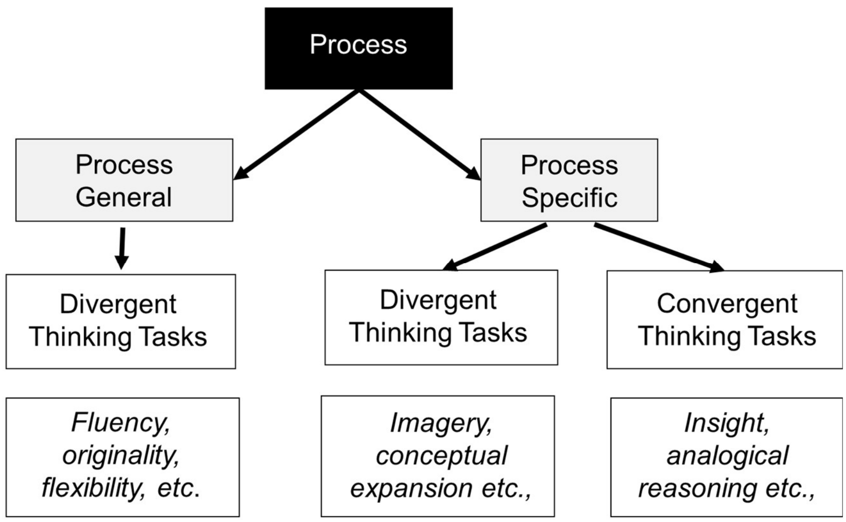
Creative Problem-solving

Creative processes have been of interest to psychologists certainly since Guildford's oft celebrated (1950) APA presidential address and indeed, before. The tracking of the moment of the genesis of a new idea is something which Dennett (2014, p. 13) describes as a "gold rush" area. The moment of creativity is both attractive and elusive. There have been many different psychometric tasks which map out what constitutes a creative person. These commonly consist of large correlations of different self- assessment measures (see for example Karwowski & Lebuda, 2016) and there has been some success in this regard. However productive this research field is, it is still unclear how far the results map on to creative action in its complexity. Ideation cannot take shape without realisation even if that realisation is only in an internal monologue. All creativity has a form. Creating is an action-driven process that unfolds over space and time, shaped by a complex array of factors and forces, many originating outside the creator; creativity is not the product of a stable set of features inherent to the creator (Glăveanu, 2020).

This is acknowledged by many in the field (see for example the 18 authors of Glaveanu et al., 2020)[97]. However, when it comes to *how* people are creative rather than the sorts of people who are creative[98], the creativity research field has encountered more challenges: The creative process is idiosyncratic and unpredictable and so does not map easily on to the laboratory tasks that are required for stable correlational analysis (Abraham, 2013). To deal with this, proxy measures for the different types of hypothesised cognitive processes have been developed (see Figure 7.1).

Figure 7.1

Measures of Process Based Creativity (adapted from Abraham, 2019, p. 43)



[97] Alongside the 18 authors of the "Manifesto", the two editors of the journal also showed support for the ideas expressed in an editorial.
[98] Or believe themselves to be creative, Creative self-belief and self-efficacy are key predictors of other psychometric measures but have only a weak or no link to the process measures of creativity (Karwowski & Lebuda, 2016; Royston & Reiter-Palmon, 2019)

These measures have been developed to trap creativity as an act (albeit one induced by the experimental situation) rather than a trait although performance on these process measures is still driven by static outcomes and often correlated with personality traits. When it is measured in the psychologist's laboratory it tends be done in one of two ways: divergent thinking tasks and convergent thinking tasks. Divergent thinking refers to the ability to generate a large number of ideas and is related to creativity because it measures the ability to generate original thought (Guildford & Hoepfner, 1971; Runco & Acar, 2012). Divergent thinking tasks focus most often on the quantity of ideas that a participant can generate. For example, a common task is the Alternative Uses Task which requires participants to generate as many uses as they can for an everyday object such as a 'brick' and are rated on fluency of idea generation and the ideas are often rated for their overall creativity either through comparison with other participants or by objective judges (Beaty & Johnson, 2020; Reiter-Palmon et al., 2019). Convergent thinking is often measured through the use of Compound Remote Associate Tasks (e.g., Bowden & Jung-Beeman, 2003) which require participants to find the link between three words (e.g. CAKE, BLUE, COTTAGE; answer CHEESE).

Other problems are recruited into the service of testing creativity with greater or lesser success. For example, Kirsh uses the word production task (Chapter Four) as a measure of creativity although it has not been used as such beforehand. The insight problems we have encountered previously (Chapters Five and Six) are also often used as a measure of creativity (Kaufmann, 2003) although this use is being challenged (Beaty et al., 2014). These tasks reflect a stage model of creativity such as the one proposed one hundred years ago by Wallas (1926). In his model of how a creative thought is generated, Wallas identified four main stages: preparation, incubation, illumination and verification. The first and last of these stages involve a conscious control (the chapter in which he makes these claims is called the “stages of control”) but incubation and illumination are beyond conscious control. It is this aspect of non-effortful thinking that makes them both of interest to cognitive scientists because they defy simple analysis and yet, so hard to examine under a field which is predicated on a linear computational model (Malafouris, 2015; Chapter Two).

This four stage model which situated creativity as a process has been reduced to a focus on the moment of illumination: This spark is seen as the moment which is of interest and which determines the success of the other processes. While it is acknowledged that leading up to this illumination stage there is a cognitive trajectory incorporating moments of both preparation and incubation, these are only understood in the light of this moment of insight. In other words, there is a form of inversion which occurs here. The moment of insight determines the quality of those prior stages rather than those stages being successful in their own right. For example, the success or not of a period of incubation is measured by the insight which is generated. Equally the process of verification is assumed to reflect analytical and thoughtful processes, but little thought is given to how this aspect forms a recursive relationship with illumination, perhaps shaping the nature of the illumination and moreover guaranteeing its success.

Such methods of assessing creativity are not neutral, they carry a hefty ontological load. Furthermore, they presuppose four things which are rarely explicitly examined. First, the most important part of the creative process is the ‘thought’ which is presupposed to be an internal cognitive mechanism—even Kirsh viewed his hints as “seeds” to thoughts (Kirsh, 2009). Second, that this thought can be consciously experienced and identified. Third, that the cognitive mechanisms can be reliably and systematically evoked by laboratory tasks and fourth that these mechanisms are reliable predictors of creativity more generally. Note that Wallas based much of his theory on the reports of Poincaré which are not entirely reliable (e.g., Weisberg, 1986)[99] so there is a tenuous empirical basis for the reliability of these predictors.

This reflects an ontological position that views creativity as reducible to cognitive processes which can be isolated both in terms of their location and their timing. Although there is hand waving in the direction of the difficulty of transferring from laboratory tasks to 'creativity' as manifested outside the lab, it substitutes for creativity a circular concept. In other words, by making the a priori decision that creativity requires certain cognitive processes and predefining the mechanisms that undergird these cognitive processes, there is little that remains to be tested. If researchers assume that creativity is the moment of insight and hunt for that, they do so without testing that initial hypothesis that creativity is linked to insight. Chapter Six discusses the problems with isolating the moment of insight and forcing homogeneity onto what the evidence is increasingly suggesting is a heterogenous process, this chapter will be discussing if this moment is linked to creativity as we commonly understand it. That is the act of creating something.

Clearly, thinking in the absence of interaction with a physical substrate, can proceed effortlessly. Impressive feats of offline cognition can be demonstrated. For example, Bilda et al. (2006) in their study of creative design found that architects performed as well in terms of consensually rated creativity when they were blindfolded as when they were allowed to sketch their ideas. This indicates that on these measures of

[99] Indeed, there is an unhelpful social representation of creativity and creative genius which makes research in this area (especially with self-reports) complex (For further discussions on the unhelpfulness of the creative genius myth generally see Montuori & Purser, 1995; Ross & Vallée-Tourangeau, 2018; Weisberg, 2010).

creative ideation engagement with the external world is not necessary and indeed, the catalogue of failures described in Chapter Six suggest that neither is it sufficient. However, research into creativity should not be led by what it is possible for artists to do in extreme circumstances but by what they actually do in their day to day practice. Once we reject the mereological fallacy (Chapter One) of a modular approach and adopt a more pragmatic stance then such tasks become demonstrations of skills rather than an investigation of creative practice.

As discussed in Chapter Three, a research programme which is committed to methodological sequestering (F. Vallée-Tourangeau & Vallée-Tourangeau, 2014) commits the fallacy of transduction (Pilgrim, 2020) – that is it assumes that data from the closed system of the laboratory can translate to the open system of the real world. In the creativity literature, the problem is often described as a problem of “scaling up”. This is, itself, a curious term. It implies that the laboratory processes suffer only because of their size not because of their shape and that indeed they should maintain the shape even while they get larger; scaled up processes being presumably isomorphic. I suggest that the problem is not so much with scaling up but the need for a reconfiguration, cognitive processes are not scaled up rather they morph in shape and size and it behoves researchers in creative cognition to map the similarities and differences rather than assume a consistent structure. In this I follow the argument laid out by Glăveanu (2020) that it is not that these tasks are not accurately reflecting cognitive processes but that the processes they map are disconnected from creative action while simultaneously and paradoxically arguing for a predictive power over these creative actions.

Alongside the ontological assumptions which permeate this work, this also reflects methodological restraints. The need for a highly powered research study underwritten by quantitative analysis requires knowledge lean problems which can be systematically and repeatedly administered, and which yield easy to measure outcomes. This is important if we wish to generate large amounts of cheap data (see Footnote 41, p 66). However, as Steffensen (2017) argues this avoids the challenge of appropriate problem selection and finding (see also Reiter-Palmon & Robinson, 2009) and also constrains problem-solving into normative outcomes subject to natural laws. Again, an expansion in size of the findings uncovered by these laboratory tasks is not is what is required but rather a change in morphology. An ethnographer of thinking can map the transformations that these changes in form require. A cognitive cartographer can map the importance of these isolated features when they are embedded in a rich and situated landscape.

There is a paucity of qualitative research in creative cognition which restricts our understanding of idiosyncratic processes (Ball & Ormerod, 2017; Neves-Pereira, 2019). Chapters Four, Five and Six have demonstrated the danger of relying on aggregated means to establish the existence or not of explanatory mechanisms. These studies assessed well delineated tasks with clear outcomes and yet still idiosyncratic, unpredicted and non-conformist behaviours were observed. If we move beyond lab-based studies of creativity and assess the rich empirical data from semi-structured interviews or case studies, these provide supporting evidence for a model in which the material environment is not accepted as a passive scaffold for creativity but actively shapes the process (Malafouris, 2014). For instance, the work of Glăveanu et al. (2013) challenges this person-centred notion of creative agency across broad range of creative domains: artists, designers, science, writers and musicians. What emerges from the results of their series of interviews is a productive tension between human and material agency (see also Ross et al., 2020) . For example, the artists interviewed suggest that “objects resist the intentions of the artist. All of a sudden, objects “ask a question” and very often “change the original plan,” being “stronger” than the creator, “imposing their rules’.” (p. 5). Designers spoke of a collaborative relationship with the materials “from the need to explore materials, to ‘test their limits’, to the frustrations one experiences when not “feeling” the fabric” (p. 7). Even the chemist (not a domain traditionally associated with material play) described their creative process in terms of a “game with matter” (p. 8)

Similar descriptions of material agency arise in the description of an artist arranging photos in Sjöholm (2014) where the final work is described as “a hybrid of the artist and the physicalities of the collected” (p. 510). The artist further “explains how the photos were *asking* to be categorized in a different way” (p. 509). Equally, the qualitative work undertaken by Glăveanu (2012) on the decoration of Romanian Easter eggs indicates that creativity is distributed between the maker and the material affordances of the eggs themselves where the affordances are initially ‘unperceived’ but will go on to shape the final product once noticed. Creative processes are constituted not by internal computations over mental representations of the artist’s materials but through and with those materials. This material engagement is what Malafouris (2020) calls ‘thinging’: ‘things actively participate in human cognitive life’[100] (p. 2). In the case of creativity, creative agency cannot be reduced to one or other part of the relationship between the maker and the material, rather creativity arises from the embodied dialogue[101] between the two.

---

[100] The difficulty in fore fronting the role of objects in cognition is driven in part by their ‘humility’ (Miller, 1987). Malafouris and Renfrew (2010, p. 1) write: “One could say that things are to human intelligence as the eye is to sight, i.e. constitutive and yet invisible”

[101] In comments on a draft version of this Vlad Glăveanu made the point that we are hampered by a language that strays close to embodying inanimate objects with attributes that they do not have. Material cannot dialogue because it does not speak nor does it make a particularly good dance partner. In personal correspondence, Paul March has made similar comments on the notion of dialoguing with something which has no voice and warns of the dangers of anthropomorphising the relationship. However, the limits of language are such that there is no way of fully encompassing the relationship between animate and inanimate actants.

Such results foreground the importance of the material in creativity and invite a more nuanced characterisation of agency than traditional linear models (Glăveanu, 2020; Glăveanu et al., 2013; Malafouris, 2014). Although there have been several calls to recentre the material in our understanding of human behaviour and cognition (Barad, 2003; Orlikowski, 2007), it is perhaps particularly surprising that the relationships with the material is under emphasised in creativity research (Tanggaard & Beghetto, 2015) where there is a necessary entanglement with material; to create is to produce something concrete, the ephemerality of a novel thought moves from imagination to creativity when it is enacted and not before. The work in this chapter, therefore, suggests that creative processes are constituted not by internal computations over mental representations of the artist's materials but through and by those materials (Bardt, 2019; March, 2019; Wheeler, 2020). This is not to suggest that internal processes are erased but rather that they are manifest through the act of creating and it is this act which is necessarily engaged with material.

The Current Study

The study reported in this chapter is not designed to showcase problem-solving of the kind which we encounter in the lab, the kind which involves this sort of disembodied and effortful thinking so beloved of the researcher in cognition. Although a problem is introduced and indeed, solved, the broader implications are uncertain. It simply posits a way of understanding creative problem-solving as idiosyncratic and hetereoscalar and embeds the conclusions from the preceding chapters in a case study of artistic creativity. There is also no value judgement. It is easy to disparage the small-scale tasks such as anagram solving addressed earlier as unreflective of "true" creativity—this is a further underlying premise to the idea of "scaling up" that some processes are somehow further up the cognitive or applied hierarchy—but it is not clear that the process documented in this chapter are any less local or parochial in nature. Certainly, it is uncertain how far the actions of sculptor in an atelier in Geneva can predict innovation in a vaccine laboratory in Oxford or how these processes might morph when resituated. The descriptions here require us to step away from a scientific practice that praises universal, non-contingent, predictive laws. As I argue in Chapters Three and Eight, a descriptive programme of science has equal validity. I further suggest that in a field which is conceptually underdetermined and lacks an empirical basis for many of its hypothesised mechanisms such a programme is necessary.

The problem-solving in this chapter is an example of creative problem-solving in action. The data were collected from midway through an artistic project by the sculptor Paul March (PM)[102]. From his perspective, the project was designed as an exploration of material engagement and there was a recursive investigation of using material to understand the experience of making and using making to understand material in practice. PM created a clay with a higher than usual fibre content. This material was specifically designed to create a challenge because of its inherent difficulty: It was an epiphenomenon of exploration and of testing boundaries**.** The work culminated in a sculptural piece called *Welcoming down the blessings* displayed as part of the exhibition Repeat in London in

---

[102] Paul March is an English artist based in Geneva. He currently works mainly with clay, which he uses to explore ambiguity and indeterminacy using primitive and strangely elegant forms. A previous career as a clinical psychologist specialising in neuropsychology left him with an enduring fascination with how the brain perceives objects. And so, as an artist he became interested in creating objects that defied categorization. In so doing he was increasingly convinced that the act of creation occurs, not in the brain but in the part of the world where the hand touches the clay. He is now using art as a tool of research to try and establish the whereabouts of creation and sensation. As part of this endeavour he is studying part-time for a DPhil (archaeology) at the University of Oxford. For more on his work see https://paul-march.com/home.html

December 2019[103] (see Figure 7.2). The exhibition consisted of over 120 flowers all made from the experimental material. The flowers have subsequently been displayed in other exhibitions.

PM and I have worked on other projects (March et al., in preparation) together and he has also worked with my supervisor (Vallée-Tourangeau & March, 2019). I did not seek him out to conduct the study rather the study arose from the research relationship we had already formed both through the collaborations outlined above and an ongoing involvement in the HANDMADE project[104]. This pre-existing relationship allowed the rapid development of a trust-based relationship with regards to the artistic process that is investigated and reported here. It also allowed me unprecedented access to the process of making and allowed me to travel to Geneva to work with PM prior to data collection for this study and subsequent to it.

The study reported in this chapter tracks how the intentionally formed problem of the recalcitrant material was transformed into the flowers above. The problem here was not a knowledge lean one which was generated by experimenters to elicit a certain behaviour but neither was it a natural problem which would reflect mundane, naturally occurring

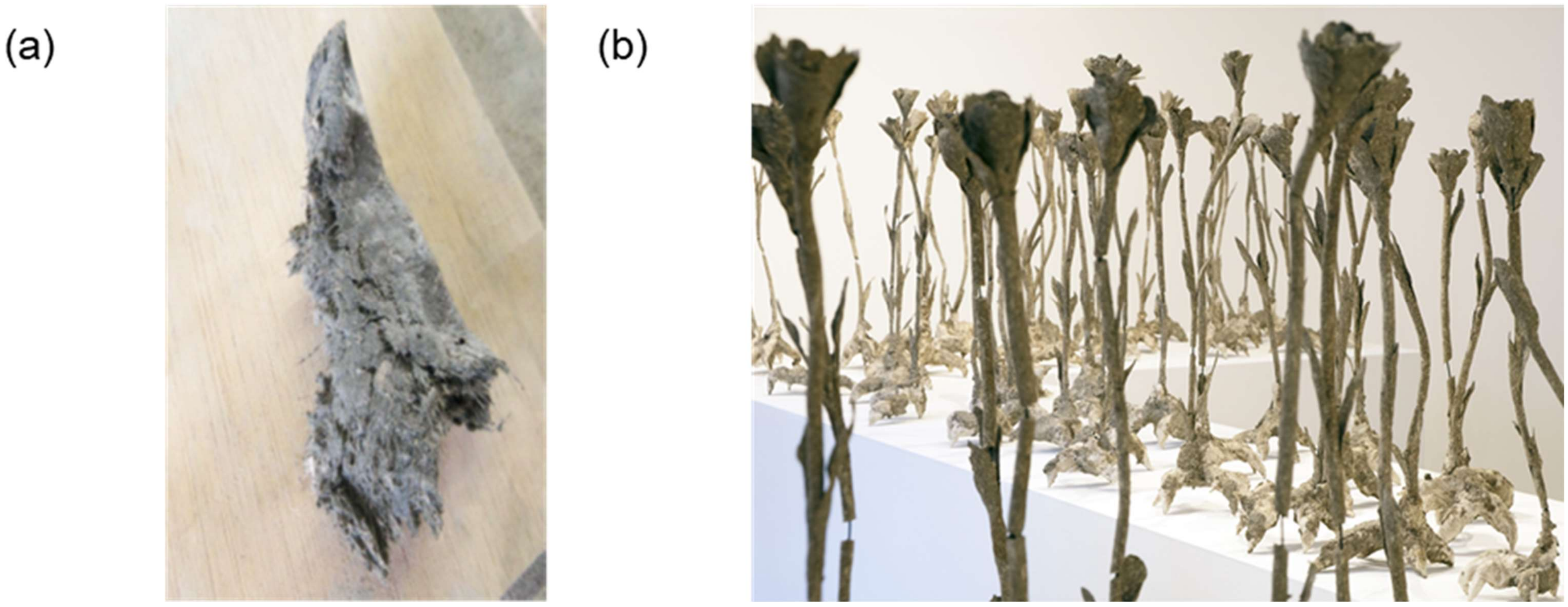
---

[103] As part of Repeat curated by Brian Kennedy for Rear Window.London: https://www.rearwindow.london/copy-of-virginia-leonard-3

[104] https://cordis.europa.eu/project/id/771997 The HANDMADE project's plan is to "study pottery making at first hand through sustained multi-sited participant observation in several traditional ceramic workshops spread around mainland Greece and the Islands. We will use a combination of methods from anthropology, archaeology and embodied cognitive science to record, measure, describe, compare and analyse the exact ways by which craft practitioners use their hands to produce a variety of material forms. We shall be collecting our data using extensive video recording, photography as well as through semi-structured interviews and interaction analysis. Our research procedure, grounded on material engagement theory, is designed to facilitate a heightened responsiveness to the details of action and the properties of the materials and the tools involved. Our broader aim is to use our knowledge about the creative entanglement of the hand and the clay and lay down the basic conceptual foundation for an archaeology of handmaking over the long term." PM and I were involved in the pilot work in February 2019 and will be involved in the project once the restrictions owing to Covid-19 are lifted.

Figure 7.2

The Transition from (a) Initial Material to (b) Final Piece.



activity The clay was a deliberate problem created in order to generate creativity:

> This time I began with a very high ratio of dry-weight fibre (50%) and then added more clay until the mixture became just about workable. The mixture began to hold together at a ratio of 10% fibre. […] With this amount of fibre in it, the material was quite unlike working with normal clay. It was not at all elastic, it tore easily and was unresponsive to the touch and tended to resist all modelling gestures In short, it thwarted any attempts to engage with it as though it was normal clay. (If This is Not Art [ITINA], 258 -276

The artistic project was intentionally formed with the background of Material Engagement Theory (MET) in mind because this is the disciplinary background of PM. Like Malafouris (2008a), PM believes that "the plasticity of clay helps to exposes the metaplasticity of the mind-material relationship in a way that is less accessible with other materials." (March, 2019, p. 141). Therefore, to understand the creative background of PM we must first review MET.

Material Engagement Theory

MET is a meta disciplinary position which has its roots in cognitive archaeology[105] but is inspired by the movements in cognitive philosophy and cognitive psychology and has subsequently inspired those disciplines. While the 4E approaches (see Chapter One) still tend feature boundaries between the physical and the material (Malafouris, 2018), MET aims to dissolve those boundaries between the object and the person and centres the 'in between'. It therefore has a far greater emphasis on the engagement with a material world (a consequence of which is what Malafouris [2015]) terms metaplasticity than other

[105] In the words of Renfrew (1993, p. 248) "Cognitive archaeology is the study of the ways of thought of past societies (and sometimes of individuals in those societies) based upon the surviving material remains." Those individuals are not available for psychological study like contemporary individuals so rather much must be inferred from the material goods which have been left. It is for this reason that Malafouris (2013) suggests that this subdiscipline of cognitive science has had an early focus on materiality.

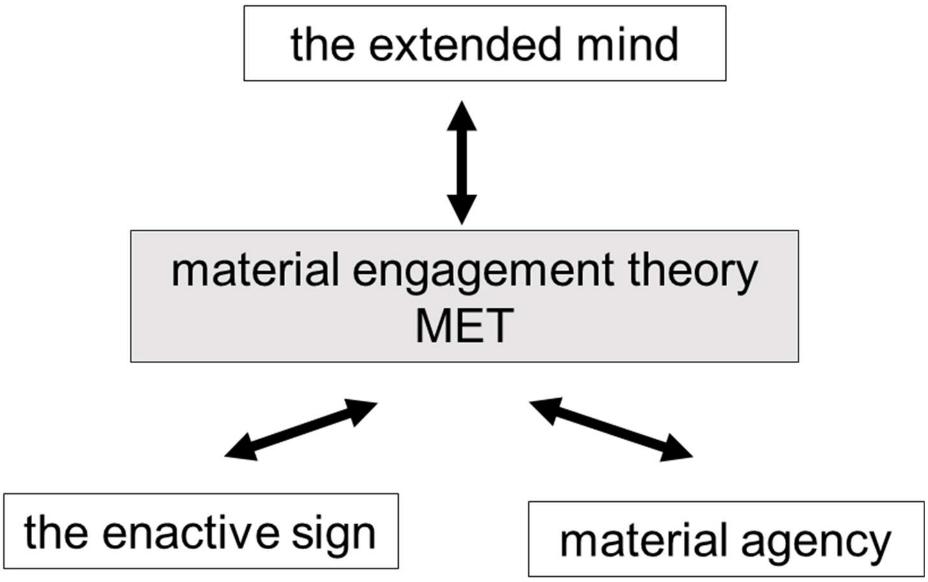
externalist traditions (Malafouris, 2018) where there is still a surprising reluctance to engage with the material. It also takes a longer term view of cognitive development that other traditions in 4E cognitive science, reflecting its disciplinary background (Malafouris, 2013). This makes it one of the more ontologically radical positions on cognition. The positioning from cognitive archaeology which has a more established tradition of taking material culture more seriously than other social sciences licenses this focus (Malafouris, 2013). MET seeks to maintain an "ontological messiness" (Malafouris, 2015, p. 352), suggesting that the solution of the appropriate level of analysis outlined in Chapter One is to have shifting boundaries in line with action. MET proceeds from several theoretical principles which I shall outline here.

Metaplasticity

The first theoretical underpinning to explore is the idea of metaplasticity. This concept underlies the argument that the material and the biological are enmeshed in such a way that they develop in a co-constitutive relationship. The theory builds on developing evidence that the brain is not a fixed organ which can only shift at certain key developmental stages but is rather in a constant process of undergoing dynamic structural alterations depending on experience and environmental input (for a review see May, 2011). Across a range of papers (e.g., Malafouris, 2008b, 2010, 2015), Malafouris has argued for a broader conception of metaplasticity to encompass "the emergent properties of the enactive constitutive intertwining between brain and culture" (Malafouris, 2013, p. 46). Here the focus is not on how the brain regions relate only to each other but also how they relate to extra-neural artefacts. Metaplasticity is therefore a conceptual bridge which links neural and cultural plasticity: "we have a plastic mind which is inseparably constituted with a plastic culture" (Malafouris, 2015, p. 360). The idea of metaplasticity and the inseparability of human cognition over time with material culture licenses the following three hypotheses of material engagement outlined most fully in (Malafouris, 2013b; see Figure 7.3).

Figure 7.3

The Nexus of Material Engagement Theory (adapted from Malafouris, 2013, p. 51)



The Hypothesis of the Extended Mind.

The hypothesis of the extended mind is more comprehensively discussed in Chapter One. Malafouris (2013) uses the example of Linear B tablets to explore how this hypothesis requires us to regard archaeological artefacts as cognitive resources which shed light not on *what* was being thought but rather *how* thinking unfolded. Malafouris suggests that what we call human cognition is ontologically dependent on both brains and things, what he describes as the "*hypothesis of the constitutive intertwining of cognition with material culture"* (p.77). Malafouris rejects the idea of isomorphism between the inner and the outer because the individual properties and features of each are important. This means that the nature and role of the material is important, leading this position to be marked by an increased focused on materiality. Rather than trying to find external processes which are functionally similar to internal processes, the entanglement of internal and external leads to cognitive assemblies and it is these compounds which deserve our attention (see in this regard the transient cognitive systems in Chapter One).

The Enactive Sign.

The semiotic dimension of culture is how culture operates as series of signs. Malafouris posits "the fallacy of the linguistic sign" (Malafouris, 2013, p. 91) which he suggests is the fallacy that an object and the word which describe that object *mean* in the same way. The idea of the enactive sign contrasts with the Saussurian position where semiosis takes place on a disengaged plane which is removed from substantive reality. Malafouris uses the example of a signature to illustrate how the material nature of the signature carries implicit meaning which is not communicated in a Saussurian analysis: The meaning of the signature is also constrained and carried through physical and material properties such as the weight of the paper and the type of ink. The material sign is independent although complementary to the linguistic sign. The material sign cannot be reduced to a linguistic property and it no longer represents but rather instantiates the concept so it can be understood in experience and action.

Material Agency.

This position states that human and material agency are entangled such that intentionality and agency are neither properties of human nor material but are rather only the property of the human-material system which arises during material engagement. Agency and intentionality on this view are cast as "open" concepts which arise in action rather than a universal and identifiable essence. “Agency and intentionality may not be innate properties of things, but they are not innate properties of humans either; they are emergent properties of material engagement”(Malafouris, 2013, p. 149).

Material Engagement and Interactivity

The studies reported before now have been inspired by the research programme in interactivity (see Chapter One). The last study shifts the focus to MET. The centering of the material in MET aligns it closely with the position held by research in interactivity although it arguably goes further in dissolving the boundaries of agency. There is still an underlying functionalism apparent in certainly the first wave of interactivity research. However, these two traditions have the greatest focus on the interaction between the material and the biological (between human and non-human actants) ignoring the skin boundary in exploring cognition and agency. The radical suggestion that we think through and with things rather than just about them is similar for both. Equally both research programmes have an emphasis on the heteroscalarity of interactions and certainly for the later interactivity theorists there is an increasing emphasis on object-thought mutualities.However, both traditions have a different attitude towards the materiality of the objects adopted. Work in laboratory problem-solving has tended to require mainly that the objects are movable and focuses on this rearrangeability as a form of cognitive offload supporting both memory and computations. Take for example, the work on word production in Chapter Four. The matter of the artefacts used is not explored although they change from physical tokens (in Maglio et al., 1999; Ross & Vallée-Tourangeau, 2020; Vallée-Tourangeau & Wrightman, 2010; Webb & Vallée-Tourangeau, 2009) to representations on a screen (in Fleming & Maglio, 2015; Kirsh, 2014; F. Vallée-

Tourangeau et al., 2020). A similar lack of concern with the material qualities is seen in work in interactivity and mathematics with numbers being presented on circular tokens (Vallée-Tourangeau, 2013), cards (Allen & Vallée-Tourangeau, 2016) or numbered 'pebbles' (Ross et al., 2020). The primary concern is that the problem representation can be manipulated. Even the more recent work which adopts a more granular approach to interactivity is still interested in the external world as a model of the problem able to be manipulated rather than as the problem itself. As we have seen in Chapter Six, when the materials move beyond external representations, interactivity is not straight forward.

This interest in the movement of the object as mapped by kinenoetic analysis and illustrated by the studies carried out in Chapters Four and Five is different to the work carried out here. In kinenoetic analysis, the problem is instrumentalised so that the object-thoughts are clearly trackable: the objects' primary function is as an external representation (Kirsh, 2009). The knowledge that the participant generates through the movement of objects mirrors the knowledge gained by the experimenter by these movements. With artistic material engagement the problem is not modelled by the artifacts or represented by them whether movable or not but rather engagement with the material creates the problem space in which problems are posed and solved and from this dialogue between material and maker, creativity is enacted. Thus, the analysis which follows here will not view engagement with the objects and artefacts as a method of problem solution but as a way of mapping a creative relationship. The movements here of the ontologically unstable object is important as a way of mapping not thoughts but rather relations. Whereas participants in laboratory assessments of problem-solving demonstrate a wide range of reactions to the materials and actions on the material shape problem-solving in a recursive manner (F. Vallée-Tourangeau et al., 2020), material engagement is not *necessary* to solve the problems (as seen in Chapter 6). The very nature of the work presented in this chapter requires engagement and so form arises from action.

Material Agency in Creativity

If we expand the creative ecosystem to encompass the objects and people who play an active role in the creative process, then we move from a linear, staged theory of creativity in which an internal plan is imposed on inert matter to a looping and reciprocal process in which agency shifts between internal to external processes. We move away from what Ingold (2010) terms hylomophism, that is the principle that form is imposed on "passive and inert" (p.92) matter. Rather, this chapter suggests that form is unstable and arises out of action, kinemorphosis. Such a model suggests that what is produced cannot be explained by a reductive process that focuses on only one or the other, but rather must take into account the relationship which arises through action. Creativity in this model is transactional and relational.

The archetypal MET example of creative material engagement is the potter at her wheel and Malafouris often uses this example to argue against a cognitivist approach to creativity because:

> I consider pottery making as a prototypical exemplar of and one of the best and diachronic models of the active mind. Not only do I see the ways of pot making as ways of thinking but I also believe that one can find few other diachronic and cross -cultural examples where all the major ingredient of the human cognitive recipe are brought forth and actualised in such an explicit and to a large extent empirically accessible manner. (Malafouris, 2008a, p. 22)

For MET, creative agency arises in action and across time and requires a detailed and focused ethnographic method which borrows from works such as that done by Malafouris himself (e.g., Malafouris, 2014) but also other work in sociology and anthropology such as that undertaken by Hutchins (1995, 2010a).

In this chapter we shall look at working with clay, but the type of sculpting undertaken is different to a potter when she works with clay. The work of PM is more akin to artwork that might be produced under a traditional hylomorphic model, that is to say it more closely resembles a piece of artwork as understood in folkloric notions of art and invites the interpretation of it as following a preconceived plan[106]. The work examined here is representational—it is clearly a flower—and unfolds over a far longer time period than the potter throwing a pot, it is therefore slower with more space for contemplation allowing for a greater degree of heteroscalarity. Furthermore, its function is different, is not designed to be used and remains vexingly fragile throughout. Its decorative status places it closer to art than the craft of a potter (March & Glăveanu, 2020)

Focused Ethnography

Ethnography is a broad research field but it traditionally requires deep immersion in an community and the fieldwork consists of understanding how the different communities understand the practices and interactions (Bartholomew & Brown, 2019). There is therefore a deep and reflexive involvement of the researcher in the community and understanding is mediated through this subjective sense. This research is time intensive and expensive. Over the past two decades a different form of ethnography has been proposed: focused ethnography. It replaces the temporally extensive data collection in traditional ethnography with a more time intensive form and the nature of the researcher is as observer rather than participant, so the subjective nature of the reporting is dampened.

---

[106] Such a model can be seen in Tinio (2013) where the "mirror model of art" is predicated on three stages: ideation, application and finishing touches. The mirror model explicit delineates these stages and gives them a linear and temporally situated form.

The focus is often shifted to singular events which are recorded with video cameras (Knoblauch et al., 2013). The participants are invited to view the subsequent recordings alongside the research team, just as the environment is selected strategically to focus on the research question, so the participants are also strategically selected as experts in the subject (Bikker et al., 2017). This licences the selective segmentation of the interview transcript to only focus on the parts which answer the research question (Jordan & Henderson, 1995) because the interviews are not used as data to generate themes but rather to elucidate on the process in the video.

In this way, it is closely aligned with the tradition of cognitive ethnography as described by Ball and Ormerod (2000) which adopts an "analysis for a purpose" approach to ethnography and data generated in this way does not need to be laced with the same level of reflexivity as traditional ethnographies nor indeed as traditional qualitative work (Ball & Ormerod, 2017). This is not to suggest a lack of meaning construction through observation[107] but that the focus of the research is not on meaning construction but on actions and behaviours which can be described and catalogued.

Social sciences tend to be interested in people and the interaction between people rather than the interaction with objects. Focused cognitive ethnographies such as Steffensen (2017) and Alač (2011) take part of their understanding of how thought-actions unfold from the interaction with others. In these instances, objects are not fore-fronted, but they are part of a dynamic object-person system. This is not the case here. There is only

[107] I would argue (although it is beyond the scope of this thesis) that all research forms construct meaning and will explore this in a little more depth in Chapter Eight.

one person[108], but the interaction of interest comes from him and the material[109]. The analysis of this video-based data is not the same as the analysis reported in the first three studies presented in the thesis. There the video analysis was used to track the material traces of process as they revealed themselves through the movements (although there was a collapse when it came to Chapter Six). In this study the material itself is ontologically unstable[110], it changes form and meaning in interaction with PM so the research here traces the transformations of this object. The actions are not traces of creativity but are creativity manifest. However, by retaining a focus on the moves and transformations of the object rather than the experience of the human agent, it aligns with the position adopted previously that attention needs to be paid to objects, to things.

The work presented here closely follows the work on Easter egg decoration reported in Glăveanu (2013). In this study, the author tracked the process of egg decoration across twenty participants who were wearing a subcam, a camera that records the process from the point of view of the participant. The study was grounded in Subjective Evidence-Based Ethnography (SEBE; Lahlou, 2011) which exploits new digital technology to better understand the experience. This development allows the triangulation of perspectives and also the description of levels of activity on a microscale (Glăveanu, 2013).

The research question is not a reflexive one which required understanding how PM constructs his making experience but an examination of how the creative process unfolds. (see Chapter Three and Chapter Eight for a longer discussion on the tension between

---

[108] I was there throughout the filming and during the interviews but the focus of the research is not on our interaction but rather on the interaction between PM and the material.

[109] Again this is not to deny the inherently social nature of the study from the internalised social element which generates dialogue with other artists alongside the social nature of the research context. All action is socially situated and constructed but the research question here does not aim to answer those questions.

[110] The letter tiles in Chapter Four and Five arguably had an ontological instability as they shifted form to make additional words. This point is explored more fully in Chapter Eight

method and methodology embodied here). The aim of this research is not to study the artist nor the artist's lifeworld but to study the artistic process. Approaching this from a pragmatic perspective, the current research study aims to collect data which are not easily quantified but does not use the nature of the data to determine the epistemological position.

Idiographic Research: Inferring Causality from Single Cases

Single case studies are rare. There is something ironic about this in the case of studies on creative insight as many of the theories are derived from case studies of individuals experiencing insight. Take for example the Gestalt view of insight which was derived from Köhler's (1925) work on apes. As is clearly described in F. Vallée-Tourangeau et al. (2020), Köhler observed apes as they realised that the way to reach a hanging banana was by moving a box. Sultan displayed classic signs of insight as understood by the current literature, but Koko's process of discovery was slower and not marked by a single moment of creative discovery that could be symbolised by a light bulb. Similarly, the basis of Wallas's influential stage theory of creativity is the reports of those such as Poincaré's, so there is no small irony that a research programme inspired by case studies would attempt to replicate these individual moments on a larger scale and repudiate evidence that derives from single cases.

Furthermore, beyond the necessity of uncovering hitherto unconsidered explanatory mechanisms that has been discussed in Chapters 4, 5 and 6, the study of creativity in practice seems to require conceptually such a granular, idiographic and non-generalisable approach. By definition, creativity is the generation of something novel (Runco & Jaeger, 2012) whether that novelty is objective or subjective (Kaufman & Beghetto, 2009). While creative cognition researchers have reduced that to a novel thought to fulfil the definition of creativity such thoughts not only need to become manifest but should also be idiosyncratic and individual. It is therefore conceptually messy to assume that creativity can be generated and measured in large, "fruit fly" populations. If a result is reproducible and generalisable, it is difficult to see to what extent it can track true novelty.

Method

The data for this study were collected and analysed using a variety of qualitative and quantitative methods. The study followed a qualitative dominant case study approach.

Data Availability

The transcripts for the dyadic interviews can be found in **Appendix I**, the transcript of If This is Not Art in **Appendix J** and the video of the flower making in **Appendix K** of the online supplementary material.

Data

The full data corpus used to inform the analysis consisted of (a) a scripted presentation given at Kingston University on 09/12/2018 entitled *If This Is Not Art* (ITINA; **Appendix J**) (b) dyadic video review sessions with PM (TR) over 4 sessions and collapsed into one transcript of 1734 lines (26,779 words; **Appendix I),** (c) published research papers, particularly March (2019) and F. Vallée-Tourangeau and March (2019). The scripted presentation and the published research papers were used prior to the data collection to inform the researcher about the nature of PM's work and theoretical positioning. The dyadic interviews were used to inform the detailed video coding and are cited in support of analysis of the data set.

Video recording

The primary data set consists of a video which is 00:37:57 minutes and transposes eye tracking data in the form of scan paths onto the video (see **Appendix K**). The video recording was taken from a point of view (POV) camera which is mounted on the Tobii Glasses 2. This allowed the recording of PM's perspective but also the tracking of his eye movement. The eye tracking data were captured with Tobii 2 Glasses with no filter applied (Raw). The scan path video was exported to ELAN for subsequent coding. It is worth noting that PM wore these glasses for 45 minutes at a time before he reported feeling uncomfortable. This is a similar amount of time reported by Krug (2018) and it may be that this is the limit for these types of glasses[111].

Post Viewing Interviews

The nature of the interviews differed slightly from traditional post viewing interviews. First, they took place eleven months after the initial recording. This was to allow the project to finish it its entirety to the display in December 2019. Second, they represented a co-creation of understanding rather than understanding being directed from the subject of investigation towards the researcher. PM was clear that this was the way he wished to work and made this statement several times in the interviews

> We don't just turn to the artists and say, "Is that right or not?" And if he says yes, that is right… And if he says no, that wasn't happening like that, that it's wrong… because it's not. (TR 1720 – 1722)

In this way they were similar to dyadic video analysis where two researchers watch a portion of video together and stop the video at salient points to discuss the work

[111] At the time of writing, these glasses have now been discontinued and Tobii 3 glasses have been launched. These are supposed to be more comfortable to wear.

(Saldaña, 2016).

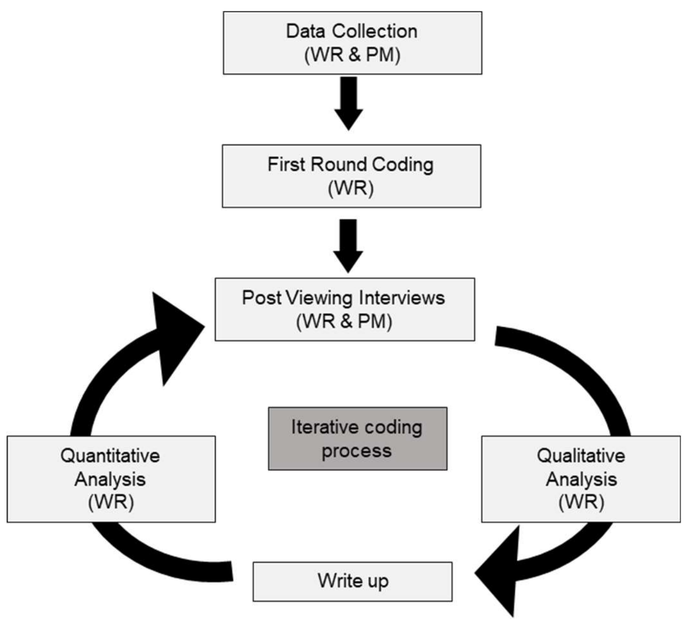
The interviews were carried out over Zoom (zoom.us) and PM and I played the video on our own devices and indicated when we would like to pause the video. The meetings were recorded and subsequently transcribed using Amazon Transcribe (aws.amazon.com/transcribe) to yield an overall text which was checked and changed in line with the interview. In total around 4 hours of interview were transcribed. As already stated, the research relationship between PM and myself had developed over several years: Significant parts of the interview did not relate to the research question explored in this chapter and were suppressed prior to the transcripts being made public.

Data Collection and Analysis

The process of data collection and analysis is illustrated in Figure 7.4. The iterative process and return to the video data closely mirror the form of GTM adopted in Chapter Six.

Figure 7.4

The Process of Data Collection from Initial Collection in June 2019 to the Final Write-Up in November 2020.



Positioning Statement

I did not approach this research as a stranger to PM nor to his artistic process. This created the benefits outlined above that I was allowed unprecedented and easy access to the creative process. I am also sympathetic to PM's theoretical background and I am aligned with him in this way. It is for this reason I avoided an interpretative analysis of the data but rather rooted the observations reported in this chapter in the video data. There are many moments during the post viewing interviews where an interpretation of behaviour was generated but this did not answer the research question and my own position in relation to PM and the process would have made it less reliable. Rather, the interviews were used to cement my understanding. The primary data source was the video file and the primary analytical technique a form of extrospection as already demonstrated in Chapters Four, Five and Six.

Results and Discussion

The video data show the creation of a flower stem with petals and leaves. Broadly, the flower required the stages illustrated in Figure 7.6, that is a stem was formed as were leaves and petals, the stem was hung from a rafter and the petals and stems were attached.

Creative Process

The initial coding of the video attempted to map the process adhering to the principle of a linear model and so the video file was broadly divided into 5 different activities (see also Figure 7.7):

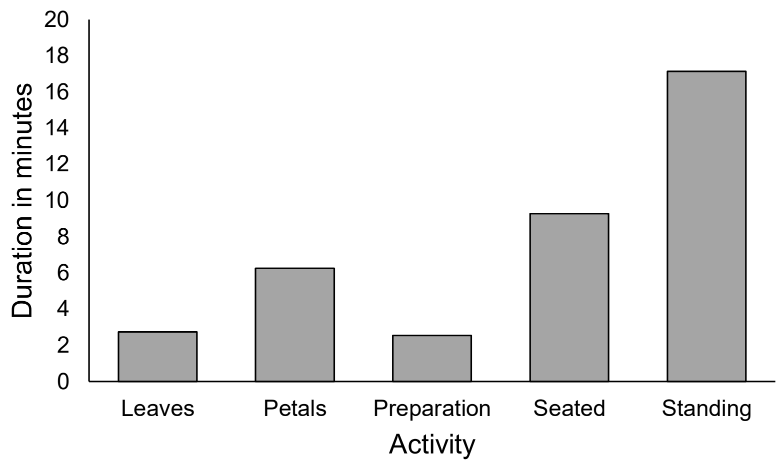
1. Preparation of the areas of activity.
2. Standing at the table (the majority of the time spent standing was spent rolling)
3. Seated (the majority of this time was spent preparing petals).

4. Placing of petals
5. Placing of leaves

The amount of time spent standing was almost double that spent sitting (17:13 minutes compared with 9:28 minutes; see Figure 7.5)

Figure 7.5

Time Spent in Making Activities



Initially, I wished to divide the time into preparation, stem formation, petal formation and amalgamating the different parts. However, the divide between the positioning and the activity is not easily segmented into ‘stem’ or ‘petal’ activities because the petals used the rolls of clay which had been discarded as too friable to be used as stem. Just as with the word production task (Chapter 4) an initial hypothetically simple binary coding scheme (in that case verifying or generating words) became trickier in practice: it was simply impossible to clearly and with principle delineate creative activities. It was impossible to say what part of the flowers was being created at any one time although the standing activities tended to be the stem and the sitting tended to be the petals. It soon became clear that the clay remains clay until it is in place and takes on a function relative to other parts of the flower prior to and while being assembled. I reflect further on this below. A systemic view resists reducing gestural processes to the end product of that process. In the end, the broad-brush categories I have outlined were all that would be easily sustained. The creative process moved between these stages in a roughly linear manner as outlined above; after all the clay started off as a lump and ended up as a flower and there is a certain constraint of order – the petals cannot be placed before they are made. However, there was also looping especially in the initial stages as illustrated by Figure 7.8..

Figure 7.6

Time Spent in Different Activities over the Course of Making the Flower

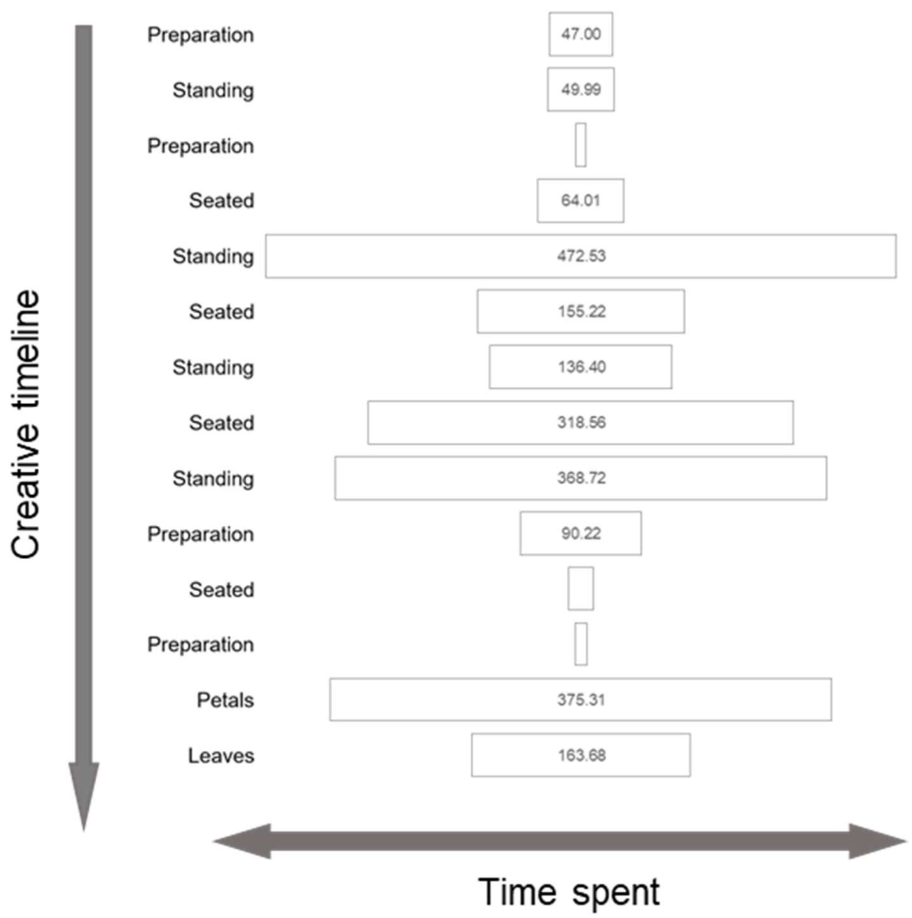


Figure 7.7

The Process of Making the Flower

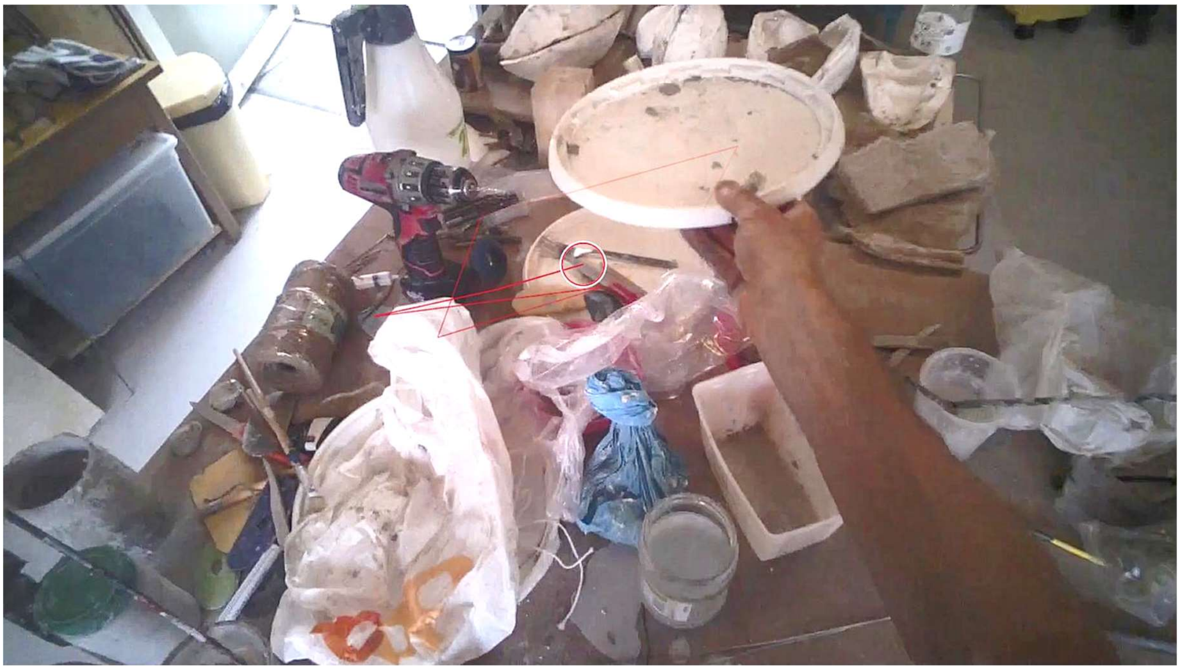
Figure 7.8

Activities Which Correspond to the Different Phases

In this figure, the time spent in each activity is not represented as a total of the overall time but is situated in a dynamic process. The dynamic and recursive nature of the process is impossible to recreate in static form (as discussed in Chapter Four) but the data on activity are transformed below to demonstrate how the activity changes over the course of the video. This illustrates more clearly that the three activities of preparing the space, standing (most associated with stem making) and sitting (most associated with petal making) did not proceed from a linear plan which was adhered to in a rigid fashion as one might expect. Note how the activity changes from standing to sitting to preparing the workspace and back. This shifting in activity is not present in the latter half of the video where the material is being assembled. This reflects several things. First the atelier which provided the stage for making the flower was messy (Figure 7.9). This mess was not the artistically inspiring environment often romanticised in the literature on workshop space (e.g., Sjöholm, 2014) but acted as a constraint on the actual process of making: "a previous untidiness is constraining my behavior and making it from… from the task point of view less efficient" (TR, 141-142). This meant that PM had to break off from tasks to create more space.

Figure 7.9

The Untidy Nature of the Worksurface. This Constined PM's Actions in an Unhelpful Way and Led to Him Changing and Shifting the Material on it.



Second, the clay itself was too wet to work with properly which led to a certain amount of anxiety and indecision: "I'm sort of just… just seem to be uncertain about what to do" (TR, 245). Humidity is the problem which dominates the first part of the creative process in the video extract. All problem-solving is the resolution of uncertainty[112](Arfini, 2019; Beghetto, 2020). The uncertainty experienced by PM at this stage is similar in kind to that experienced by participants in the studies reported in Chapters Four, Five and Six. It is how this uncertainty is resolved that constitutes the micro-level problem-solving. The difference in the case is that the uncertainty is not generated by an artificial situation and nor does it then have a guaranteed correct answer. However, in all cases answer to how the

[112] Uncertainty or ignorance. The two are different in nature but closely related.

problem is solved is revealed by the detailed scrutiny of the video data.

Creative Problem-solving

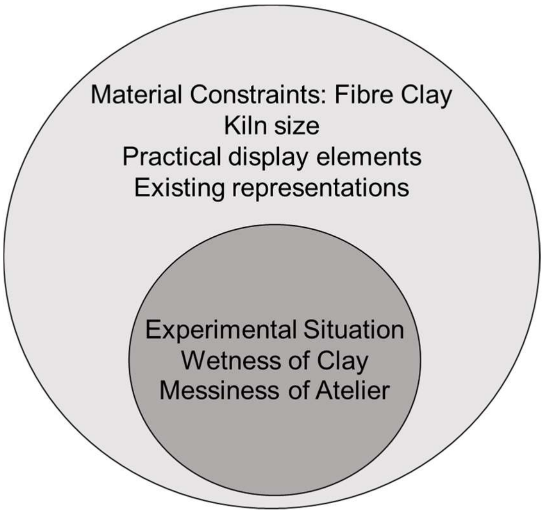
Creative Constraints

Creativity arises through constraints (Glăveanu, 2020) which can be intentionally generated or arise in the normal course of working through and with material. For example, the fibre-clay was an intentional constraint but the characteristics which permeated that intentional constraint were not predicted. Through the analysis of the data corpus, two different levels of constraints were identified which were initially divided into two types: conceptual constraints and procedural constraints. These are illustrated in Figure 7.10.

Conceptual constraints were those which influenced the higher-level elements of the project. The planning stage if we adhere to a model such as that mapped out above which view art as moving from concept to material form of that concept. These took on several forms with more or less mundane implications, but they can be broadly conceived of as constraints on the overall design of the project or an initialization phase (Mace & Ward, 2002; Tinio, 2013).

Figure 7.10

The Two Types of Constraints: Procedural (Inner Circle) and Conceptual (Outer Circle)



In this were included the unusual clay-fibre mix which constrained the type of sculpture that could be created, the size of the kiln (see below on stem formation for a more detailed description) and also the existing socio-cultural prescription on what constitutes a flower. Although PM was not intending to represent a flower as is commonly understood through the idea of representational art and has created pieces of art which deliberately frustrate this idea of straightforward meaning (March, 2019), the expectations of the parts of a flower – the petals, the stem and the roots constrain the development of the final piece. The text from *If This is Not Art* make this uncertainty about how to make the constituent parts into art explicit and situates this becoming through reference to other sculptors (Vanessa Hogge; https://vanessahogge.com/, and Phoebe Cummings; http://www.phoebecummings.com/). It is a flower only because of the attendant pre-existing expectations and the social representations[113]. The size of the flower was also constrained by the size of the kiln. The clay was too delicate to be easily transported so it had to be fired in the kiln in the atelier which meant that the whole flower needed to fit into a 60cm kiln. All of these things constrained the design of the process. Although they operated on different levels from socio-cultural expectations to very practical constraints, they have in common that they pre-existed and influenced the conceptual design of the flower. There was a polytemporal collapse of place, instruments, culture, and expertise.

The procedural constraints arise from tensions in the creative process itself as the act unfolds. These fall more into the Expansion and Adaptation phase as described by Tinio (2013) in which the original plan has to change in line with problems encountered. Recall that this was a repeated process which developed along similar procedural lines each time. As is narrated above, there were several moments within the process documented on this day which constrained the production of this particular flower. First, the messiness of the atelier which restricted the rolling space for drying and meant a much more recursive process of kneading and collapsing the clay was required. Second, the moisture levels of the clay. Clay is particularly sensitive to different levels of moisture and needs to be neither too wet nor too dry. The clay on this particular day was too wet. Finally, behaviours and processes were changed by the experimental situation: "There were two tasks for me. It normally there's just one task and it… and then… and in fact, probably a main task was doing the eye tracking because that was so unusual." (TR, 29-30).

What is particularly interesting is how many of the conceptual constraints filtered

[113] This is an example of the necessary socially situated nature of practice. In Chapter Eight I shall discuss the underlying nature of socially situated practice and how it underlies an understanding of the object across all three studies.

down to procedural level and how many procedural constraints filter up. For example, this initial categorisation places the fibre-clay material as a conceptual constraint, but the nature of the material constrained the very procedure employed which in turn led to the concept of a flower. The conceptual constraints are enacted through the material which is dependent on a process of handling. The concept only exists in a form which is generated by a process. Again, attempts at sharp binaries were frustrated. Perhaps the strongest conclusion we can draw is that such concept and procedure resist separation into a linear planned process.

The Humidity Problem

Alongside the higher-level problem of how to engage with the material, the problem which is being solved over the first part of this video recording is the battle with the humidity level of the clay. A clay which is too damp does not respond to the artist's manipulations in a satisfactory way. What is perhaps most important to note is that this problem was not solved through contemplation and indeed may not have been solved in the correct way at all. PM did not consider the wetness of the clay, rather he enacted the same moves that he always did. In the dyadic interviews PM suggests that his process of solving the problem, that is to try to make the parts of the flower was "a mistake" (TR, 637). On watching the video, he suggests that correct thing to do would have been to roll the clay up again and knead it remove the moisture. However, despite the mistake the flower was able to be produced using the 'wrong' strategy. The moisture mostly noticeably affected the building of the stem which was resolved by 00:07:03.

This has interesting implications for the way we conceive of problem-solving. If a correct or at least satisfactory outcome can be generated by an incorrect or less than satisfactory process, then the relationship between process and outcome becomes less clear. This then has important implications for the use of correct solution as a proxy for process. Especially in open problems, sub-optimal solutions can still be solutions.[114] However, there is also a sense through which the unplanned actions also solve the problem because each action removes some of the wetness from the clay:

> My uncertainty is that again, this is a source of sensational humidity of just you know, this is humid, this is humid, this is humid. And I, should I carry on, should I carry on. And… but all these gestures are at the same time taking some humidity out because any… any contact, the clay has something else it removes some of the moisture. (TR, 254-257)

So, the state of uncertainty is resolved not through contemplation but through action. The actions in this case solve the problem without a conscious plan.

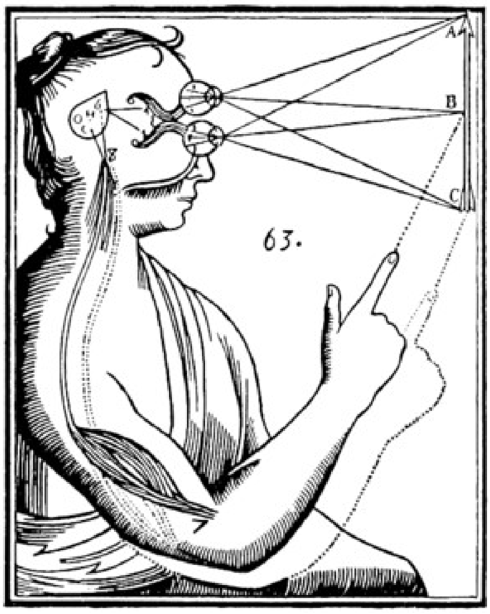
There is an important reason for the link between action and the removal of uncertainty that we see here; the perception of moisture occurs in a haptic rather than a visual modality. It is not possible to perceive through the visual modality whether the clay carries the right level of humidity to be manipulated, nor is it even possible through passive touch alone. Sitting back and contemplating the problem would not have yielded the necessary information to solve the problem. The manipulation of the material at once gives vital information about the state of the problem and acts as its own problem solution.

The primary focus of perception research in psychology is the visual modality. There is the oft encountered dualism to this focus. Consider Descartes' model (Figure 7.11):

---

[114] Indeed, this may be part of why the participants struggled with the socks problem in Chapter 6. If the problem is how to draw a pair of socks, 5 draws may not be the most efficient but it is certainly more efficient than selecting 5 and then returning to the bag to extract 3 socks if the ultimate aim to locate a pair.

Figure 7.11
Descartes' Theory of Vision from 1644 Principles of Philosophy



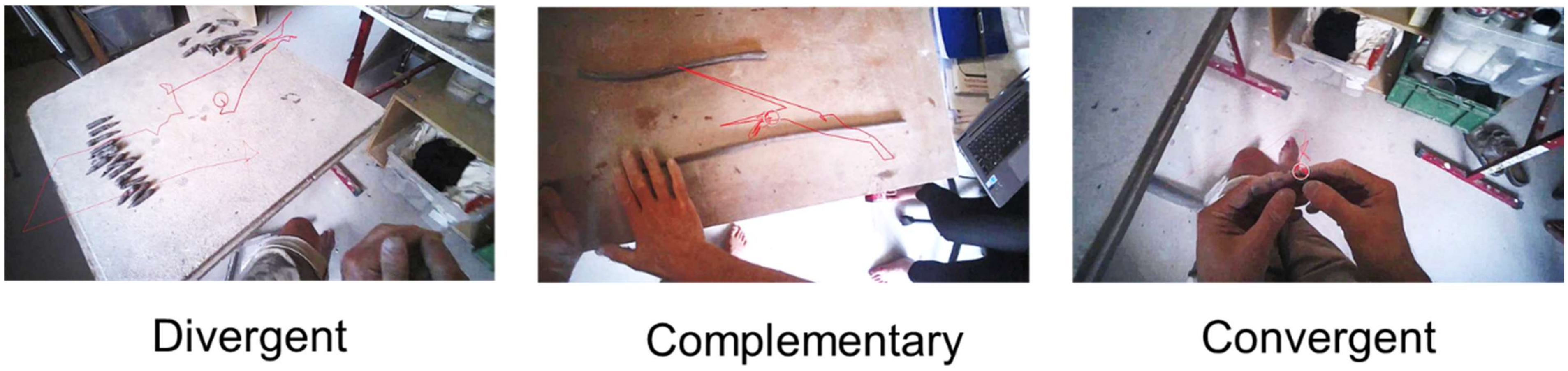
Input from the environment enters through the eyes and is converted to gestural action. Vision is input, action is output. Not much has been updated from this model in current models of perception which are saturated by ocularcentrism (Pallasmaa, 2012) to such an extent it is unnoticed. It is curious in this regard that an extended approach to perception often blinds the metaphorical perceiver and presents us with the thought experiment of a blind man walking with his stick to ask us to examine where perception starts and ends in this instance (Malafouris, 2008b; Merleau-Ponty, 1945)

However, non-visual perception does not need to only occur when visual perception is hampered or in some way constrained. Indeed, the eye tracking paths traced on the POV camera indicates how often visual and haptic attention diverged and that different perceptual modalities were recruited for different problems. Much of the information that was needed to solve the problem simply could not be perceived in the visual modality. As Ratcliffe (2018) makes clear, touch is the only causal sense, that is to say it is the only sense which changes what it is sensing. It therefore collapses the distinction between perception and action far more succinctly than the visual modality (see for example Noë [2006] for an enactivist approach to perception based on vision).

Three types of eye/hand moves could be easily detected: Convergent, divergent and complementary (see Figure 7.12). The convergent eye gazes were most obvious while the petals were being fanned. This is a delicate operation that could lead to tearing. The tearing could not be detected by touch and required visual concentration. At the opposite end of the scale when PM was kneading the clay, an automatic gesture, his eyes were never on the clay or his hands, rather they moved across the display and away. Here it is tempting to follow the eye gaze as an indicator of attention. However, at these points the primary problem to be solved was still the moisture levels of the clay and the only way of solving it and of knowing that it was solved was through the manipulation of the clay therefore I suggest that the primary perceptual input at this point was haptic rather than visual. Finally, the complementary gaze was most often seen in the act of rolling. The tube of clay was most likely to tear at one end as the rolling occurred at another because the act of rolling would cause the end to wave more uncontrollably. When rolling occurred, the hands would deal with one end of the clay and the eyes with the other.

Figure 7.12

The Three Types of Haptic Visual Perception Crossover: Divergent, Complementary and Convergent (the Red Line Indicates the ScanPath, the Red Dot the Focus of Attention)



Collapsing Pragmatic and Epistemic Actions Through Touch. Kirsch and Maglio (1994) introduced the idea of pragmatic and epistemic actions based on their research with Tetris players (see Chapter Three). The information yielded by the moves in the Tetris playing was visual and so there were moves which did not progress the player nearer a pragmatic goal but did change the epistemic state of the player by externalising the information processing. The data presented here question the clear binary between the two types of actions. Rather if we switch the primary perceptual instrument to haptic rather than visual, then the pragmatic actions also held an epistemic value. Rolling the clay was a pragmatic action but it also allowed PM to update his understanding of the levels of moisture, so it yielded epistemic value. The divide theorised by Kirsh and Maglio only holds if actions are considered haptic and epistemic information is considered as visual. In other words, the divide between pragmatic and epistemic echoes the old binary that sees the world of the intellect dominated by a visual sense and the world of action dominated by a tactile sense (Pallasmaa, 2012). The data in this chapter collapses this easy distinction.

The Stem Problem

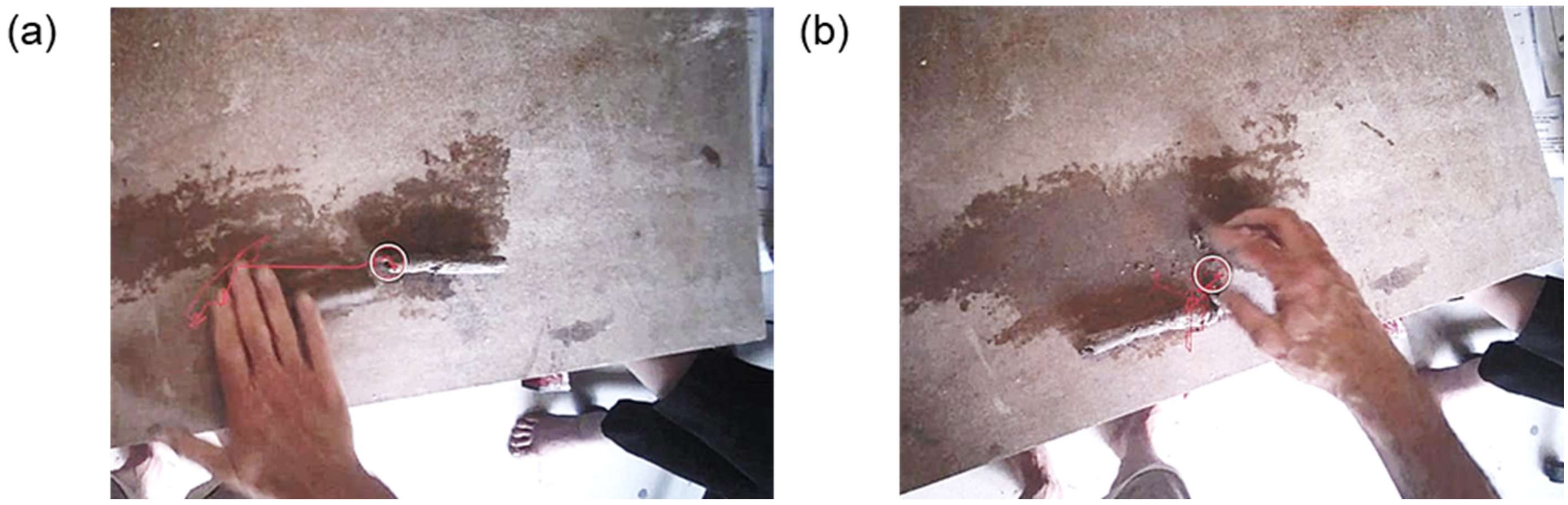
The building of the stem caused problems across different time scales. First, from a design perspective, the stem arose from the pre-existing petals. Any model of creative design needs to take into account the recursive relationship between the different elements. A stem was necessary because a petal had been made and the petals already made exigencies on the nature of the stem – "I've got the petals but they need a stem so and then its constrained by the petals what the next stage is" (PM, TR, 34-35).The form of the flower was therefore constrained on multiple levels – societal expectations and the already existing pieces. The plan, if it can be called such, to make a flower emerged from process and existing representations.

The size of the stem was further constrained by the size of the kiln – the fragile nature of the clay mix precluded using a kiln on a different site. The friable nature of the clay also meant that the stem needed support. PM first tried with string, but this was not strong enough so later a certain sort of kiln safe wire was used. However, even the wire was sensitive and revealed unexpected behaviours. PM lives in France so initially bought the kiln safe wire from France. The initial version appeared to work until it had been fired when it was too stiff and snapped. The initial affordances changed through firing and the problem solution. Finally, a wire was sourced with the right properties. The problem here of creating a stem which would hold the petals was determined by the properties of the pre-existing sculptural elements and the nature of the material plus a wider sociocultural conception of what petals 'need'. It is worth noting that PM thought he had solved the problem of the stem when he used the French wire, it was only through enacting the problem solution that what had seemed like a good property – the stiff nature of the wire – could be revealed as not useful in practice and this not until after firing. The uncertainty of an outcome is revealed in practice.

Over the course of the primary data source, the video extract, it was the stem which caused the most pressing original problem and also caused the most tension between PM and the material. The stem was formed by rolling the clay into a tube, the action of which required standing at a table. The act of rolling caused problems. The clay would break or crumble over the course of rolling or it would stick to the table or PM's hands (see Figure 7.13).

Figure 7.13

The Roll of Clay (a) Breaking Mid Roll or (b) Crumbling Mid Roll. (Note the Wet Surface)



It also caused anxiety with PM checking frequently and regular breakages to the stem. Once the rolling had produced a successful stem at 00:07.03 when he leaves it, the number of ruptures drop substantially (aside from the introduction of the micrometer which is discussed at a later point).

The material properties of the clay also dictated a change from the rolling technique that would be traditionally employed in ceramics. It was not until the dyadic interview that PM noticed that he was not using a two-handed rolling technique: “And that's a slight mystery to me” (TR, 403). The reason that he developed a one-handed technique was because the nature of the material meant that when it rolled with two hands the stretching technique which was part of the rolling would break the fragile clay.

> It just means that the two hands aren’t pulling. If you roll like that normal that normally what you do, you’re rolling out, you also stretching the clay. But if I stretched this clay it just breaks. I'm just saying this because if an experienced ceramicist saw this without knowing the mixture of the clay they would think he doesn't know what he's doing and umm..[…] It's to do with a particular and…. and is directly connected with the quality of the material which it goes against, uh, which…. so I've actually being taught a different way of doing this, (TR, 412-422).

In this case, the nature of a skilled engagement with the material required PM to change his actions.

The problem of the stem is not solved through a moment of creative insight. Indeed, the problem is solved so gradually it is hard to pinpoint the moment when it becomes clear that the roll is going to be successful. At 00:06:30 PM selects the wire to gauge the length, which is the first sign of hope, perhaps this would be the time to suggest the roll has become successful but note that it needs to be verified. At 00:07:03 the stabilized length of the stem has been achieved. The point to make here is that the problem solution is not a crucial event pivot as it might be in a laboratory-based study. Perhaps the kneading earlier solved the problem, perhaps just time and the heat of the day. Some problems resolve themselves over time—a material incubation. The second point is that the solution of this problem was only part of the overall creative arc. It was not lingered over but attention swiftly moved on. Creation – that is the turning of toilet paper, hemp and clay into a flower – is systemic, can only be understood as that. There is no one moment where the flower becomes a flower.

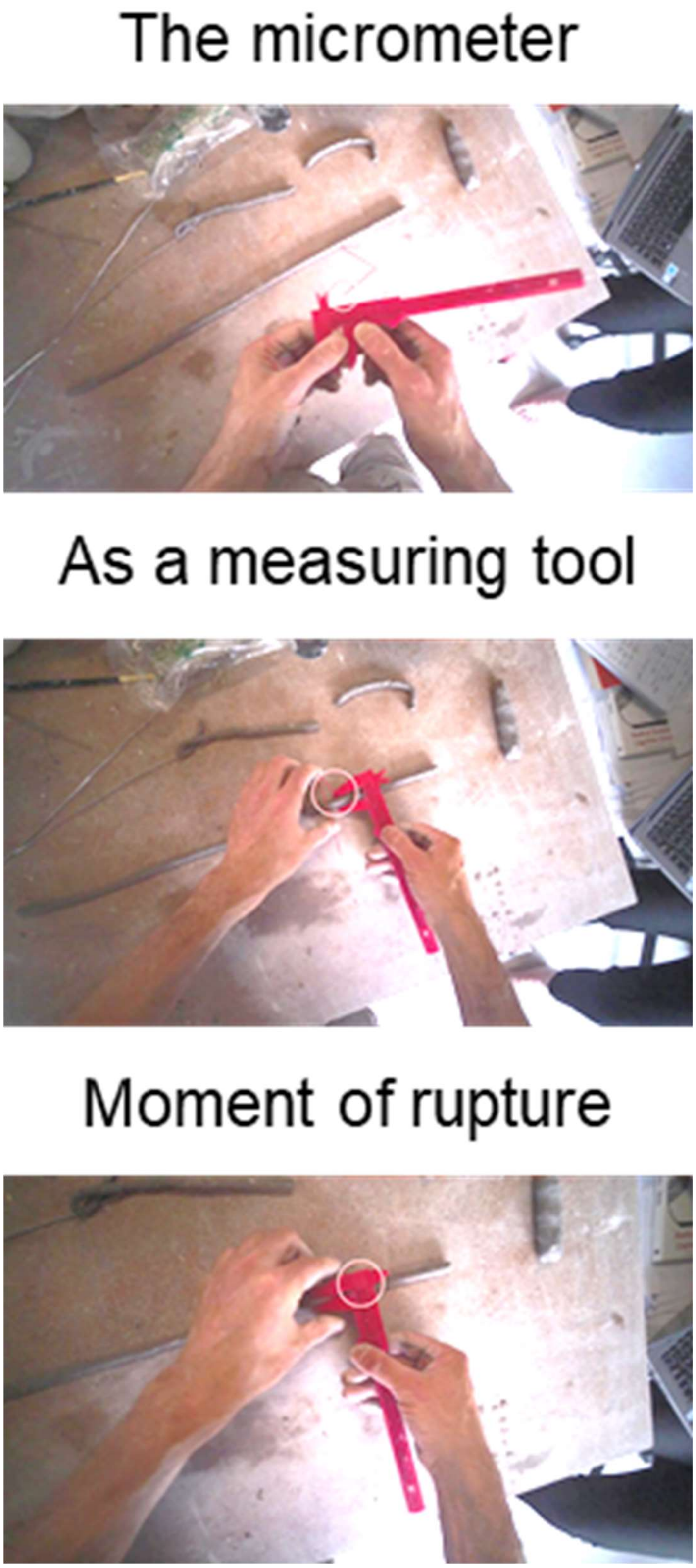
Locus of Knowledge: The Problem of Repetition

Part of the problem to be solved here also lies in repetition. The whole set of flowers that constituted *Bringing down the blessings* were formed in a similar way and yet were also unique. The conceptual framework and the procedural framework converged to reproduce similar forms. The problem then became how to repeat the knowledge with a regularity of form. The location of the knowledge of how to reproduce those forms, how to solve the problem of repetition in idiosyncratic artwork was solved in three ways, two successful, one less so. The first was the use of a micrometer at 00:07:58 (Figure 7.14).

It is not clear what the function of the micrometer was. The stem width was checked and noted but no attempt was made to alter the stem width. Indeed, the measurement has no function in the system: “I have a memory that I don't know what the measurement should be” (TR, 490). The incongruity of the micrometer is enhanced when it slips and breaks the system.

Figure 7.14

The Use of a Micrometer



A second way of ensuring the regularity of form was the use of different objects as markers of sorts. The clear example was the piece of wire cut to 57cm which was used to measure the length of the stem. There were various constraining factors on the stem length both as a function of the delicate material which could not support additional weight, the size of the kiln and the capacity of the material to hold its form either in the stem or in size congruent petals. Petals also functioned as boundary objects for each other. Once PM had formed the first few his eye gaze showed that he used the existing petals as guides to structure and constrain the size and shape of the ones he was currently making.

The Problem of Definition: Existence in Function

The ontology of a piece of clay is determined in part by its purpose in the flower construction: ontology and function are inseparable. There is nothing which makes a piece of clay anything other than a piece of clay beyond its existence as *art*. There are moments when a petal is created or a leaf when they turn from clay to flower piece. This moment is contingent and different for each observer. This is the interplay between perspectives and affordance. When perspectives lead action, they use whatever affordances there are which allow the realisation of the concept but the, as the work proceeds, unexpected affordances emerge and perspectives need to be realigned (Glăveanu, personal communication).

The material was not intended to make flowers, rather it was resistant to other possibilities. Instead, during the course of "fiddling" with it, PM created something he recognised as a petal.

> I returned to my workshop and began the now familiar gesture of pressing the clay between by thumb and forefinger. But this time the experience has a different signification. Although I am making the same gesture that I have made many times before - this time I see it taking the form in front of me of a damaged, desiccated petal (ITINA, 475 - 478)

This fiddling is a clear example of an exaptative action. That is, it started as a gesture without a clear plan which developed into an intentional act both in this initial moment of "outsight" but also over the course of the project as a whole. Further constraints shaped the ongoing creation. The flowers required a stem[115]which had to be strong enough to support the head of the flower, yet the material was fragile and did not lend itself to a stem of that sort.

The flower made in the video analysed here was not the first time that the flower was formed. From the viewpoint of creativity outlined above, there is nothing "novel" or "useful" in the actions. A creative cognition approach would search for the moment that the 'idea' to make a flower arose but as the documenting of the mundane process will reveal, that simply is not the right question. Creative ideas cannot be disentangled from their manifestation: manifestation and ideation become one and the same thing.

This is not just a question of semantics but reflects the process of making with a material such as clay. This indeterminacy both generates and constrains the form. This can be most clearly seen in the indeterminacy of the rolls of clay in the first six minutes of the video. Three rolls are formed. The first roll is formed but is too friable to serve as a stem. A second is formed but it is also too weak. PM uses the first to create a series of petals. This is not a systemic approach – at another time he pulls lumps of clay off the clump to

[115] And a root but the root is not addressed in this chapter because at the time of data collection it remained an unsolved problem.

form into petals. Indeed, it may be that the timing of this video in the middle of the two stages of experimentation and “production line” captures the evolutionary change:

> What I found out is that rolling out the stem in one bit is quite difficult, it often breaks. And so I have spent, you know, two minutes, which seems a long time rolling out and then it would break […] Then I'd have to start again. And then I realized that I could use... But it was in doing this ..I realized I could use these bits of stem to, use the petals. […] But what's interesting from our point of view is that it's not a hard thing. I.. It happened over a number of sessions, but slowly dawns on me it. It really is a.. a sort. It's difficult to know at some point I'm obviously…I’m conscious of it now, but I think I was doing it, but I really even knowing that, I realized this. (TR, 104 -112)

This reads like a moment of microserendipity (see Chapters Four and Five) not occasioned by a fortuitous accident but rather by something going wrong. Subsequent to converting the first roll, PM rolls a third which breaks in rolling. He kneads this but discards it. There are no visual clues that would allow an observer to understand what led to the choice or discarding and yet it was this moment that decided whether that part of clay would become a petal or a stem.

Conclusion

The work presented in this chapter is a tentative way of approaching creativity outside of a psychologist’s laboratory but employing the microgenetic techniques which have already been explored earlier in the laboratory tasks reported in Chapters Four, Five and Six. The data here collapses several existing binaries between form and function, procedural and conceptual constraints and action and perception. The analysis of the data presented here borrowed from the kinenoetic analysis developed in Chapters Five and Six but developed it to look not only at knowledge derived from action but also form derived from action: kinenoesis turns to kinemorphosis.

What is also evidenced here is the mundanity and heteroscalarity of the creative process. The idea that the moment of insight so sought after in the laboratory should be the moment which requires trapping like some magic pixie dust to fully understand creativity does not map onto the evidence presented here. There was no one moment in which the clay turned from clay to artwork, rather the moment of creativity was temporally as well as situationally distributed. The flowers in the final exhibition were created many times. The flower that we see created here is not the same flower that was displayed in London, at that stage it had gathered more meaning, more companions, a root, in short it had become art. This suggests the emphasis should be less on the elusive cognition process but on the actions extended through time and across people and things which coalesce to create a new creative object or event.

To conclude, I wish to argue for a focus on process. The work reported in this chapter highlights mechanism and behaviours yet to be considered in creativity, it does this by unveiling creativity. opening up and revealing something which was the property of a process rather than a person or an object. I argue that focusing on how the person and product emerge as creative from the process and locating creativity there can help shed light on some of the complexities of that remain yet unexplained. The creator becomes creative through process and the material becomes a creative material through that same process. *Doing* is what is creative, not *being*.

Chapter Eight : Overall Discussion

The work presented in this thesis is wide ranging. It starts with a philosophical analysis, takes in experimental studies conducted in the laboratory and ends on a detailed and focused cognitive ethnography. The data are quantitative and qualitative and the analytical techniques move from statistical analysis to qualitative, grounded theory analysis. The tasks employed also vary; I look at word games, verbal riddles and an artistic product. However, across both the theoretical and empirical and the plurality of analysis, there are a few simple recurring themes. This triangulation of theory and empirical evidence is reassuring although I will offer some critical caveats to this. This chapter will synthesise these themes alongside offering some limitations to the research presented here. Finally, I will suggest potential profitable avenues for future research exploration.

Shifting Perpsective: From Where to How

In Chapter One, I surveyed the philosophical arguments relating to the locus of cognition. This was a necessarily brief discussion and the recap here shall be no less brief. In short, those who cluster under the umbrella of 4E cognition reject the notion that the locus of thinking should be brain bound:

> Thoughts do not occur in the brain, but in one's study, in the library, or as one walks down the street. The location of the event of a person's thinking a certain thought is the place where the person is when that thought occurs to him. Thoughts are to be found written down in texts but not in the head of a human beings (Bennett & Hacker, 2003, pp. 179–180)

For proponents of these positions, cognition should be studied as emergent from a system of people and things. In contrast, a more classical approach treats cognition as internal and modular. I propose that rather than concerns about the fixed locus of cognition, it is more appropriate to consider thought as polymorphous and resistant to simple categorisations as either inside or outside the head. It can clearly occur with no engagement with the world: It can occur in sleep with no conscious awareness (as in incubation, see for example Gilhooly, 2019) or in effortful but immobilised processing (for example Aizawa's [2017] locked in syndrome example from Chapter Three). However, it also arises in cognitive coordination with objects and people (see for example Steffensen, 2017). My simple argument is the latter is both more prevalent and more importantly, easier to track in well operationalised task environments. The theoretical focus on internal cognition has excluded the engaged form from serious consideration. The transactional nature of thought too often excluded from consideration.

The introductory chapter to this thesis situated itself within the tradition of interactivity but, more importantly, adopted the pragmatist approach of watching what people *do* in a pluralistic blend of inductive and deductive analysis. The data presented throughout the course of this thesis suggest that thought happens in and out of the head. They are not intended to establish a research programme which works against internally directed cognition but rather to query its primacy and to set up a parallel investigation into how actions over objects generate thinking. My own theoretical stance is sympathetic to the externalist position; however, I argue more strongly that a dogged adherence to a theoretical position can limit research opportunities. Chapter Six clearly demonstrates the requirement to return to action and behaviour as way of refining our understanding of how cognition happens. Theoretically plausible and attractive propositions do not always transfer from the abstract to the concrete. The benefit of the methods employed in this thesis are that they unveil some thoughts but I do not seek to anchor all thought in any one place be it inside or outside the head. The focus is no longer where cognition is located but how it unfolds, process supersedes locus.

To be “in thought” is often contrasted with being “in action” but the data in this thesis suggest that this binary is an unhelpful one. Thought and action were entwined throughout in ways that were hard to separate (see Chapter Four). In short, I suggest that to be “in thought” in a sequestered and nonactive way is not only difficult to evoke (see Chapter Six and the fallacy of the unsullied lab below) but also not *necessary*. An initial aim of the thesis—to test empirically whether or not the environment was recruited in thought—upheld the anecdotal observations that when there is the opportunity, thought is scaffolded by action and that action occurs with and through objects. Chapter Four established the additive benefit of action and Chapter Five, its ubiquity. At times actions fell outside what could easily be coded or traced and coding of a more qualitative nature ensued as in Chapter Six and Seven. However, the kinenoetic field could still be mapped through actions on the world and things outside the brain. Given this is the case, then the greatest puzzle may be why a research programme has developed to focus on the mysterious and rare brain bound thoughts when action-thoughts are visible and traceable.

The Nature of Intentionality and Goal Directed Action in an Open System

At the start of Chapter One, I outlined the importance of problem-solving for the study of cognition by citing Dixon et al. (2014, p. 160) suggesting that a key issue for psychology was“ ‘How does an organism successfully engage in goal-directed action?’”. While still holding that goal-directed action is important, the evidence from this thesis casts doubt on whether all action is necessarily goal-directed or rather whether that goal is necessarily a cognitive one. Rather it adds a third form of action to the dyad of epistemic and pragmatic action suggested by Kirsh and Maglio (1994) by suggesting that we can consider actions which do not proceed from a plan: exaptative actions (see Chapter Six and Chapter Seven)

The current models of extended cognition and interactivity are rooted in an agent centric view of thinking with the environment in which an intentional agent actively recruits the external world to support their thinking. The thought is still directed from a plan which is presumably internally, if unconsciously, generated. The data here suggest that when agent engages in a strategic use of the environment then the expansion of the cognitive workspace much as predicted by SysTM (G. Vallée- Tourangeau & Vallée-Tourangeau; see Chapter One) does support problem-solving and cognition. Chapter Four showed an additive relationship between the time spent interacting with an object and also suggested that a high interactivity environment scaffolded the use of a more efficient strategy. Chapter Five assessed the role of chance and found that simple accident is not enough to generate solutions and that a model in which the agent was passive did not augment success. In other words, agency and strategy play key roles and this should not be underplayed.

However, the fully agent centric model of interactivity as a form of cognitive offload and expansion of the cognitive workspace requires an omniscient problem solver (see Chapter Six), that is one who already knows the best pathway to solution and is able to recruit appropriate resources to scaffold that knowledge. In these cases, the agent-centric model of interactivity comfortably explains the increased performance. On this view, we recruit objects around us to support problem-solving much as a computer recruits more RAM and or a cyborg might decide on corneal implants to better play Tetris (cf Clark & Chalmers, 1998). The data in this thesis broadly support this view. Chapter Four suggests a scaffolding effect for interactivity, Chapter Five suggests that unplanned changes are not enough to restructure problems and Chapter Six points to a complex system but one where interactivity in the semantic field scaffolded success. There is little evidence presented here to overthrow the agent in problem-solving tasks in the laboratory merely to augment that view.

Things become more complex when we move outside the laboratory into a different problem-solving or cognitive environment. Chapter Seven does just that and just as "mental" tasks (the solving of riddles) necessarily require mental activity, the creation of a material object naturally unfolds through action. Creativity is thought made manifest and the articulation of creativity requires a form which is often a material one[116]. It is therefore not surprising that creative cognition occurred in this case over and through materials and, furthermore, that intentional actions were hard to pinpoint. Chapter Seven describes a mix of heteroscalar thought processes which resulted in a creative product which compressed multiple and distributed intentionalities into a single point in time. When the processes that went into this product were expanded, they revealed an emergent agency in which both objects and person are necessary in way that they are not in the sort of mind games commonly explored in the psychologist's laboratory.

Additionally, actions over objects change the topography of the problem state such that the actions themselves generate novel goals so goal-directed and intentional actions can also be goal generating and intention producing. The linear model is collapsed and action and intentional goals are co-produced. As noted above, I have labelled these actions which are weighed with an intentional fluidity as "exaptative" borrowing from an evolutionary term which describes a repurposing of original intent. These refer to actions

[116] A creative thought is not creative until it is articulated. Creativity requires manifestation even if it is only as internal dialogue. It is not possible to be unconsciously creative. Most creativity requires then a physical form and it is often in the generating of this form that the thought is changed.

which change purpose when in line with the changed topography of their making. This can refer to the changing of an end goal state from an initial one – in the anagram task the movements can have the initial (impossible) intention of making MUSIC out of SINUM and the moves reveal MINSU and the goal swiftly changes to MINUS – but they can also refer to actions which entirely change their initial purpose. For example, fiddling with the clay (Chapter Seven) bought about the starting point of a petal which then changed the action. Such actions are typically left out of a programme which seeks intentionality (indeed I deliberately excluded them in Chapter Four and Five) and so their importance remains to be seen.

The Importance of Materiality and The Object

Accompanying the shift from the locus of cognition as being contained within the thinking agent, the work in this thesis suggests that a serious consideration of the external environment requires a parallel consideration of the nature of the object. This focus on the object aligns the methods employed more closely with Material Engagement Theory (MET; see Chapter Seven) but it also has two key differences I would like to highlight here. First, and most important, the temporal scope here is far more limited than MET. The work presented in this thesis focuses not on how thinking developed from an archaeological perspective of symbiotic co-development but on how a thought emerges from the shifting ontologies of objects, actions and epistemic space. Second, while there is an engagement with ethnographic method, the disciplinary basis for the work in this thesis is experimental cognitive psychology and so there is a reliance on experimental techniques and a plurality of analysis which mixes quantitative and qualitative work. I suggest that this plurality of disciplinary perspective will strengthen our understanding of materially engaged cognition.

The research reported here was only made possible by a systematic coding and observation of the movement of objects through the kinenoetic space. The detailed and granular method of analysis developed from the experimental manipulation associated with interactivity, that is allowing participants to interact with a movable problem representation. It is thus more detailed and less reliant on phenomenological report than MET. In this way, it aims to situate itself with the broader cognitive psychology field. For example, there is a growing literature on gesture and its role in cognition (e.g., Alibali et al., 2011; Goldin-Meadow & Beilock, 2010; Hostetter, 2011; Hostetter & Alibali, 2019) and it seems likely that such a literature could be complemented and extended by a focus on the object as not only a method of tracking thought but as a constituent part of thought itself. A move from embodied cognition to an interactive and enacted one which requires taking the material world and its objects seriously.

There was an effect of material in the research reported here. Chapter Four conceptually replicated Kirsh (2014) where letters were presented on a computer screen and shuffled with the press of a button. In the study in this thesis the digital letters were transformed into physical tiles and the shuffle condition required participants to pick up the whole set of tiles. The research demonstrated that the time and cognitive cost required by moving from a digital to material interface erased the benefit of the shuffle condition. Indeed the data suggested that it is not shuffling or moving which is important but the relatively fluidity of each, the speed of shuffling in Kirsh (2014) more closely mirrored by the fast moving letter formations of P41, actions over materials that would be enacted differently in a point and click environment. It seems plausible that the very mundane nature of the sock problems in Chapter Six led to the participants' inability to mix problem spaces from a household task to a riddle to be solved in the abstracted space. Finally, the material was centred in Chapter Seven when the clay-human system was explored. Research which requires engagement with the material world is required to engage with the properties of that world.

Object-Thought Mutualities

In an agent centric model of extended cognition, the object is neutral and is important only in terms of its functions. When the object is centred with the agent in cognition then it takes on an equal importance. A kinenoetic programme suggests that the ontology of the object shifts as the epistemic state of the agent shifts and that the two co-constitute each other. A cognitive object (and by cognitive, I mean anything which is part of a cognitive system) changes its ontology even if its form remains static.

Different objects were involved in the research presented in this thesis and they shifted differently through action as they became part of a cognitive ecosystem. This shifting of ontological status is clearest in the case of the clay in Chapter Seven which over the course of the video changes from a recalcitrant material to a flower ready for firing, form and ontology both changing through action. However, it is also the same for the other objects in the thesis. Take for example the letter tiles employed in the studies in Chapters Four and Five. Their material form remained the same throughout the experiment. However, in movement the ontological status of the string was changed and these movement-induced changes rendered the object cognitive. Static letter representations are not cognitive, no processing occurs, cognition happens as the letters flow through a series of planned and unplanned actions. The objects are ontologically reconfigured to become traces of thoughts and thinking. In so doing, they further altered the cognitive ecosystem and the thoughts generated so that the objects were both shaped by thought and shaped it: Object-thought mutuality.

In Chapter Six, the objects resisted changing their ontology in line with the experimental situation. Their mundane nature and the opportunity they offered to make concrete concepts which were only abstracted in the second order version of the problem restrained participants from engaging fully in the experimental situations. This stubborn ontology shifted and changed the participants thoughts. In this they refused to become cognitive objects. In Chapter Six, I discuss the nature of the objects employed in interactivity research and there is perhaps a value in a level of abstraction when dealing with artificial situations for the objects to become cognitive. In this study, the cognitive objects which most structured the problem-solving in this task were the instructions. First I had to reconfigure these as the researcher to support the co-construction of the experimental situation and then they were also reconstructed by individual participants as their understanding of the instructions developed along the pathway to solutions. The instructions reconfiguration was driven by shifts in epistemic state but also caused those changes so thought and cognitive object were co-constructed.

Laboratary Experiments in Open Systems

The work in this thesis charts an uncomfortable and somewhat wavering course through the implications and contradictions of experiments on open cognitive systems. It is unclear if it always succeeds, however, I will make the strong claim that all cognitive systems are necessarily open ones. In light of this, I suggest it is important to reflect on and consider the shape of an experiment.

The Fallacy of the Unsullied Lab

It is often stated that embedded and situated cognition will better explain "real world" problem-solving. Such a position is not neutral: it sets up a clear binary between the "real world" and the unreal world of psychological science which deals in controlled and modularizable entities. Classical research in problem-solving and traditional research in interactivity rely on the idea that an experimental condition free from environmental complexity can plausibly be generated. For more traditional research, this means that the external world can be ignored and for research in interactivity this allows the unreal world of the laboratory to be contrasted with the real world outside of the laboratory and to be found lacking.

The parameters and characteristics of what constitutes a real world are rarely explicitly delineated, it is simply enough that it is "other" to the experimental procedure. Thus, the real world can be the classroom, the artist's atelier, the office…. anywhere other than the laboratory. It suffices merely that the "real world" displays complex and uncontrolled explanatory mechanisms. In contrast then, work in the laboratory depends on creating a controlled environment so that the manipulated variables can be accorded causal status. In both externalist and internalist conceptions of experimental psychology, the laboratory is a clean space unsullied by the complexities of the "real world"..

There are two problems that I have outlined with this position over the course of this thesis. First, such a position adheres to a modularity of mind which I suggest commits the mereological fallacy. It requires us to accept that the processes in this controlled environment are the same as those in a complex environment and so underlying nomological principles can be extracted and generalised. Chapter Four demonstrated that different strategies were enacted differently and other research which takes in process (Bocanegra et al, 2019; G. Vallée-Tourangeau et al, 2015) also suggests that cognition is different in different environments. Therefore, sequestered cognition may only reflect sequestered cognition and the idea that the environment can be added later misjudged. Second, it also requires us to accept that the unreal and unsullied world of the laboratory is sustainable and that experimental controls can be applied so casual relationships can be established in an uncomplicated way. This thesis has demonstrated that adherence to the key manipulations is variable (Chapters Four, Five and Six) and cannot be assumed. Furthermore, Chapter Six demonstrated the difficulty in even maintaining a controlled environment and suggested that the quantitative data did not reflect the problem-solving that was occurring in this task. All psychological research is necessarily situated and ignoring the situation is to ignore the necessary complexity of cognition.

Objectivity and Subjectivity in Data and Analysis

The primary data source for each of these studies was video recordings. These data naturally lend themselves to a plurality of analytical techniques and the use of video invites an almost unbounded granularity of analysis which is demonstrated in this work presented here. Across the quantitative analysis of outcome measures generated through experimental manipulation (Chapters Four, Five and Six) and the systematic coding of behaviour both inductively and deductively (Chapters Four and Five) to inductive coding of behaviour to generate more substantive theories in the tradition of Grounded Theory Method (Chapters Six and Seven) these data have been treated equally in terms of their empirical import and also their objectivity. No one type of data has been approached as more objective than another. This section unpicks this epistemological stance.

My own theoretical position has been clear from the start and is reflected in the experimental situations I designed, and the performance measures employed whether quantitative or qualitative. This level of theoretical subjectivity runs through the data reported here. Chapters One, Two and Three suggest that there is a contingency to the measurement of cognition. First in the selection of the focal centre of the locus of cognition, second in the types of tasks employed to investigate thinking and third in the outcome measures. The selection of these reveals the researcher's theoretical biases and restricts the research field in a subjective way which is perhaps underexamined when it comes to quantitative research. These biases are no less evident in the selections that I made and the decision to focus on behavioural markers and search for explanatory mechanisms beyond those which are already hypothesised. These decisions reflected theoretical allegiances.

The experimental situation is an artificial and constructed one and the behaviours produced are also constructed by the situation and the understanding of the researcher. There is a contingent relationship between the researcher's understanding and the understanding of the participant and rather than a subjective/objective divide, a form of intersubjectivity arises where both participant and researcher co-construct their understanding of the experimental situation. This understanding is internally coherent and upheld by epistemological circles which reinforce the validity of the situation. Chapter Six is perhaps an excellent example of what happens when the co-construction of the experimental situation is derailed.

Two of the studies in the thesis used an insight rating scale which relies on the participants having not only the same understanding of their own experience as each other but also as the researcher. The underexamined problems are clear, for example in Chapter

Five. In this study, I was interested in whether a participant encountered impasse in problem-solving. Two formal measures were employed –an in-trial report and a post-trial report –and these were triangulated with observed behaviour. That there was no clear overlap between any of these three measures suggests that there was no clear and stable underlying concept. Participants clearly displayed moments of impasse in the eyes of an observer without recognising (or at least reporting) which indicates a disconnect in the co-construction of experimental understanding. The participant's experience of impasse is constructed by the researcher and by the participant both in and out of the moment. It is somewhat ironic that the only directly observable, and therefore verifiable, act is the one that falls into a traditionally interpretative epistemological bracket, whereas reports generated by the participant do not.

This is not to say that such behaviours and actions were not also constructed and rooted in a social situation. Watching the videos of participants struggling with the sock problem in Chapter Six, it is notable how many of them held their head in their hands or assumed postures similar to Rodin's *The Thinker* to indicate their mental state. The richness of video data licenses a series of analytical approaches, however, uncovering a hidden mental state through an interpretivist paradigm was not the aim of the research reported here and the data were not analysed in this way. Rather the mental state of interest was that which was revealed in and through the movement of objects. These movements were directly observable and verifiable. Mental states made visible rather than inferred either through language or adherence to hypothesised models.

Additionally, the data employed here were not linguistic reconstructions of events. Transcribed language featured in Chapter Six and Chapter Seven but was not analysed as data in itself rather as illustrations of a process. There were opportunities to unpick how a participant constructs their own reality, but this was not the aim of the work produced here. Such an analysis would return the focus to the head, the internal construction of a thought mediated through language. The work in this thesis aims to push outside of the head and again return the focus to what people do rather than what they construct.[117] Therefore, in line with classic Grounded Theory Method, the data collected in this thesis were treated equally whether classed as quantitative or qualitative.

The analytical technique was almost recursive and interpretative from the point of data collection. As the researcher, my first contact with the data was while engaging in data collection. The nature of the experiments conducted here and the design meant that I was observing participants while they were participating in the research as well as through deeper and more detailed video analysis after the event combined with quantitative analysis to understand deeper trends. The analysis I employed was a dynamic, organic process that led me from qualitative to quantitative and back again and rendering traditional dichotomies unproductive. I experienced a deep familiarity with the data while collecting it before re-experiencing it as a video and as performance outcomes on a spreadsheet; the analysis in this thesis emerges from these perspectives. In Footnote 42 (p. 67), I discuss how the experimental situation privileges the theoretical post hoc interpretation of data rather than the in situ work which is often taken on by those who are part of the design or the interpretation whether that be a research assistant or even computer software through internet mediated research. This distance from the data reduces the opportunity for spontaneous observation of those mechanisms which can arise outside

[117] People do also language outside the head and language is equally a cognitive tool (see Love, 2004; Wetherell, 2007) but it is the material objects that are of primary interest in this thesis. It is plausible that such objects constitute a further semiotic field which should be explored independently of people's understanding of them.

of parameters decided a priori.

Limitations of the Research

There are several limitations of the research in this thesis. I shall outline them here before moving on to discuss ways in which the research programme can be extended to investigate these limitations further. Rejecting a modular theory of cognition brings with it its own complications. The work here does not clearly solve the level of analysis problem laid out in Chapter One, nor is it clear on the use of the experimental method discussed in Chapter Three, and finally the process tracing which may help us solve some of the problem alluded to in Chapter Two is also unsatisfactory.

The Level of Analysis

Traditional methods of assessing embedded or extended cognition are still rooted in a cognitive and computational model of mind and contrast low and high interactivity environments to demonstrate an augmentative effect of interactivity. The research in this thesis suggests that the reality of experimenting in an open system is more complex than this supposes. Let us return therefore to the initial question posed in Chapter One which discussed the appropriate level of analysis when conducting research in an open system. It is unclear that the work in this thesis has made any progress on where to draw the boundaries. For example, Chapter Four consciously excluded actions over objects which had no effect on the final representation and focused only on the letter arrays and there was no focus on other forms of interactivity whereas Chapter Six was open to any information which structured my understanding as a researcher of what was driving solution rates: One was deductive-inductive-deductive focusing only on easily operationalised behaviours whereas the other one was inductive when the initial deductive logic failed to explain the behaviour. In other words, there was no consistent level of analysis.

Second, the laboratory experiments reported in this study (and to a lesser extent the study reported in Chapter Seven) made the conscious choice to restrict the level of analysis to a temporally boundaried moment. While there was a horizontal extension and a push against the traditional limits of an experiment in terms of explanatory factors and mechanisms, these factors were only those that were directly relevant and observable but they do not extend back in time causally. Related to this, the participants were treated as epistemically neutral and interchangeable from the moment of entering the laboratory (thus while there is an emphasis throughout on the fallacy of the unsullied lab, the work in this thesis still adheres to a possibly more pernicious fallacy, that of the interchangeable participant). This became more obviously unsustainable as the research progressed. Assuming that the participants were interchangeable experimental units led to problems in the pilot studies in Chapter Five where common English words were unrecognised by some and in Chapter Six where the same set of instructions were understood in different ways so that the original notion of a binary epistemic state of familiarity or unfamiliarity with the problem was not adequate to profile the differences between the participants. Current problem-solving research exercises strict controls over stimuli and instructions to ensure a consistent epistemic state across participants but it may be that a more nuanced profiling is necessary even before the experiment takes place to ensure consistent epistemic state.

Process Tracing

While Chapters Four and Five use lettered tiles to track efficiency and proximity to a novel answer, it is not clear that this is as clear cut as is claimed in these chapters. The tracking of proximity relies on a form of lexical realism[118] which suggests that the letters hold a semantic value outside of the system formed by the letters and the meaning held in the head of the problem solver. This was apparent in the pilot study for the anagram task: the selection of anagrams was not sensitive to the prior knowledge of the solvers and so word combinations were produced that were not recognised as such by the participants. The word SCOUR was a clear example of this and several times participants in the pilot task expressed disbelief that this was a real word. This motivated the familiarity task in the main study in this chapter (see Chapter 5). The letters in themselves hold no meaning without being incorporated into a system of letters and person. This means that movement and action shift the participant's epistemic field in a nonlinear manner so that proximity to an answer is not easy to measure. However, one of the benefits of this method of kinenoetic analysis is the chance to trace process and further thought and refinements are necessary to fully exploit the potential of this method.

The Nature of the Experiment

The research reported here is ambivalent with regards to the nature of an experiment in psychology and, indeed, follows the literature by neglecting somewhat this aspect, Chapters Four, Five and Six all reported traditional experiments although Chapter Six described the limitations of this approach. Chapter Seven was not a laboratory based experiment as commonly understood by the research literature but was an experiment in the broader, Baconian definition as defined by both Hacking (1982) and Radder (2003) which takes as its definition an interference in the direction of nature. In Chapter Seven, a

[118] I am grateful for an anonymous reviewer of the manuscript based on Chapter Five for these reflections which I extend here.

tension was artificially generated (the fibre-clay) in order to stimulate creativity. The research reported here depends on a successful manipulation of the experimental parameters. Conclusions are drawn on the basis of the aggregate quantitative results that while explored further in the qualitative analysis are broadly accepted.

However, the work here also reports consistent failures of the experimental situation and the use of coarse performance measure to accurately reflect problem-solving process or even solution accuracy. Therefore, there is an unresolved contradiction. The thesis does not answer how we might use the experiments when some of the foundational assumptions have been undermined but nor does it suggest an alternative.

This ambivalence reflects an uncertainty on my part. Interfering in nature allows us to create situations in which we can measure and record behaviour of interest. Something which is bounded in time and with clear outcome measures can draw out additional and important explanatory mechanisms. I suggest that embracing the fallacy of the unsullied lab might be the answer to the problem posed here. This requires reconfiguring the ontology of the experiment from a hypothetico-deductive tool to an inductive arena where observational tools can be used to track process. However, we return to the level of analysis problem outlined above in which the selection of salient points of interest are necessarily subjective. Therefore, it may be that quantitative and experimental researchers should become more comfortable with the fabilist nature of knowledge.

Further Research Directions

The research reported in this thesis argues for a methodological change in the study of problem-solving in open systems and, from there, more broadly to cognition. I make the argument for a qualitative approach and for a dynamic and recursive analytical technique which moves from quantitative to qualitative and so disregards traditional disciplinary boundaries. This argument is necessarily in the early stages of development and there are several key areas which require further exploration and support. The main focus of this thesis has been the development of a method which can support a plurality of analysis while accounting for cognition as it occurs outside of the brain and through objects: Kinenoetic analysis. The suggestions for future directions will explore this novel method and suggest two further applications.

Kinenoetic Analysis

The method of kinenoetic analysis has yet to be fully explored. The current development of it is somewhat inelegant and evolved to answer research questions and leverage the experimental manipulation of interactivity rather than as a pre-established method. It is similar in style to the method used in F. Vallée-Tourangeau et al (2020) but that method was developed in tandem with the research reported in this thesis[119]. A grounded qualitative style of analysis which is not constructivist in nature, but which recognises the constructivist nature of all scientific discourse was necessary to fully understand object-though mutualities but this method is far from refined. However, the observations in this thesis suggest that there is the potential to unveil aspects of situated cognition which have remained hitherto underattended.

Chapter Six demonstrated the limitations of the method and the difficulty of sustaining a focus on object tracing alongside the limits of the direct transfer of second order problem-solving to a first order problem-solving environment. The problem space

[119] I am a corresponding author on the paper and developed the analytical technique alongside F. Vallée-Tourangeau.

does not only consist of the objects that populate it but is also culturally constituted parameters. The experiment in Chapter Six showed that taking a laboratory task without fully considering the nature of these parameters leads to confusion and highlights the nonsensical nature of much of the narrative lying behind mental riddles.

There are two areas of research which I suggest would particularly benefit from the application of KA.

Creativity

Insight problems are often seen as a special form of problem-solving because of their relationship to creativity. It is true that the research areas of creativity and problem-solving are very closely related – in a review of studies between 2003 and 2012 over half the experimental studies in creativity used problem-solving as a measure (Long, 2014). Additionally, insight problem-solving is sometimes directly applied to creativity as a measure of the moment of insight described in Wallas's (1926) four stage theory. The moment of insight is presumed to be a moment when a new idea is born. Each chapter has involved tools and measures which assess creativity. The word production task in Chapter Four replicates a task from Kirsh (2014) which features in a paper called *The importance of chance and interactivity in creativity*, Chapters Five and Six feature insight tasks which are designed to capture the genesis of a novel thought. Chapter Seven situates itself firmly in the creativity literature drawing from Material Engagement Theory to offer a view of creativity which is underexplored in the current research literature.

I argue therefore that all creativity as we understand it requires problem-solving, but it also requires other key processes and characteristics. It is unclear therefore to what extent therefore, the research in this thesis can add to research on creativity. Creativity is a complex and multifaceted phenomenon which draws on proximal and distal causes and actions. However, the work in this thesis demonstrates the importance of an engagement with materials in generating novel thoughts. As I report in Chapter Seven, this focus on actions on and through material is becoming an increasingly important aspect of research in creative practice (Glăveanu & Beghetto, 2020). I suggest that the methodological approaches laid out over the course of this thesis (that is a mixture of quantitative and qualitative work rooted in ethnographic case studies and microgenetic analysis of video data) are essential to understanding the creative process and I hold that a theoretical position that takes into account engagement with materials and people as well as what emanates from the talent and experiences of the creator is the only way we can understand creativity as a materially engaged form. Creativity cannot arise in a closed system.

However, this is not to suggest that I see creativity as a subset of problem-solving. Rather, problem-solving as it is operationalized in the laboratory is a special and idiosyncratic subset of creativity more generally. After all, creativity is the generation of something both novel and useful (Runco & Jaeger, 2012) and the solution of any sort of problem requires the agent to generate something that they did not know before (even if that is something as mundane as the answer to 375 multiplied by 450) as the solution to the problem. All problem-solving is a form of creativity.

Microserendipity

The use of KA allowed us to turn the experience of serendipity as seen through the lens of personal narrative into bounded event: microserendipity. This has broader implications. Serendipity is an often reported mechanism in discovery and innovation (Campanario, 1996; Yaqub, 2018), but it is refracted through narrative and personal understanding. This means that the underlying mechanisms are harder to unpick. When serendipity is turned from an experience into an event then it becomes clearer what both precipitating and inhibitory factors may be. For example, by profiling the environment and establishing moments when it yielded the solution, it soon became clear that this event was a contingent one and that generating information in the environment was not enough to generate serendipitous moments.

Rather like the limitations with process tracing above, there seems to be no simple way to establish what turns an answer in the environment into an answer for the cognitive ecosystem. There are several things that could generate this moment. There has been a theoretical exploration of the factors in the environment which encourage serendipity (Björneborn, 2017) but until the development of KA it has been difficult to systematically and objectively test them. This would be possible and would have important implications for design (Olma, 2016) Current research on the personality traits which are most likely to experience serendipity have led to inconclusive results (McCay-Peet et al., 2015). The research in this project would reframe the analysis of the prepared mind to a state (perhaps something akin to incubation in creativity and problem solving) rather than a trait and manipulate this "prepared" state to establish whether this could lead people to take greater advantage of luck in the environment (see Henok et al., 2018).

Most important to serendipity, is the "act of noticing" (Rubin et al., 2011) which appears to be related to inattentional blindness (Simons & Chabris, 1999) and irrelevance processing (Agnoli et al., 2015). This act of noticing is an underexplored aspect of serendipity and the combination of KA with eye tracking would enable the pinpointing of the moment when the accident is noticed and acted upon. This could also help clarify the problems with process tracing outlined above. In short, the method of KA allows us to explore what has hitherto remained unmapped. It seems likely that there will be profitable discoveries.

References

Abraham, A. (2013). The promises and perils of the neuroscience of creativity. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00246

Abraham, A. (2019). *The neuroscience of creativity*. Cambridge University Press.

Agnoli, S., Franchin, L., Rubaltelli, E., & Corazza, G. E. (2015). An eye-tracking analysis of irrelevance processing as moderator of openness and creative performance. *Creativity Research Journal*, *27*(2), 125–132. https://doi.org/10.1080/10400419.2015.1030304

Alač, M. (2011). *Handling digital brains: A laboratory study of multimodal semiotic interaction in the age of computers*. MIT Press.

Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological Science*, *22*(9), 1138–1144. https://doi.org/10.1177/0956797611417722

Allen, M., & Vallée-Tourangeau, F. (2016). Interactivity defuses the impact of mathematics anxiety in primary school children. *International Journal of Science and Mathematics Education*, *14*(8), 1553–1566. https://doi.org/10.1007/s10763-015-9659-9

Andersen, F., Anjum, R. L., & Rocca, E. (2019). Philosophical bias is the one bias that science cannot avoid. *ELife*, *8*. https://doi.org/10.7554/eLife.44929

Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive brain*. MIT Press.

Anderson, M., & Magruder, J. (2017). *Split-sample strategies for avoiding false discoveries*. National Bureau of Economic Research. https://doi.org/10.3386/w23544

Andriani, P., Ali, A., & Mastrogiorgio, M. (2017). Measuring exaptation and its impact on innovation, search, and problem Solving. *Organization Science*, *28*(2), 320–338. https://doi.org/10.1287/orsc.2017.1116

Anjum, R. L., Copeland, S., & Rocca, E. (2020). *Rethinking causality, complexity and evidence for the unique patient a CauseHealth resource for healthcare professionals and the clinical encounter*. Springer Open.

Arfini, S. (2019). Situated ignorance: The distribution and extension of ignorance in cognitive niches. *Synthese*. https://doi.org/10.1007/s11229-019-02328-0

Arfini, S., Bertolotti, T., & Magnani, L. (2018). The antinomies of serendipity: How to cognitively frame serendipity for scientific discoveries. *Topoi*. https://doi.org/10.1007/s11245-018-9571-3

Ash, I. K., Cushen, P. J., & Wiley, J. (2009). Obstacles in investigating the role of restructuring in insightful problem solving. *The Journal of Problem Solving*, *2*(2). https://doi.org/10.7771/1932-6246.1056

Ash, I. K., & Wiley, J. (2006). The nature of restructuring in insight: An individual-differences approach. *Psychonomic Bulletin & Review*, *13*(1), 66–73. https://doi.org/10.3758/BF03193814

Bakeman, R., & Gottman, J. (2009). *Observing interaction: An introduction to sequential analysis*. Cambridge University Press.

Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. Cambridge University Press.

Ball, L. J., & Litchfield, D. (2013). Interactivity and embodied cues in problem solving, learning and insight: Further contributions to a "Theory of Hints". In S.J Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain* (pp. 232–239). Springer.

Ball, L. J., & Ormerod, T. C. (2000). Putting ethnography to work: The case for a cognitive ethnography of design. *International Journal of Human-Computer Studies*, *53*(1), 147–168. https://doi.org/10.1006/ijhc.2000.0372

Ball, L. J., & Ormerod, T. C. (2017). Cognitive psychology. In C. Willig & W. Stainton Rogers (Eds.), *The Sage handbook of qualitative research in psychology..* (pp. 572–588). SAGE.

Barad, K. (2003). Posthumanist performativity: Toward an understanding of how matter comes to matter. *Signs: Journal of Women in Culture and Society*, *28*(3), 801–831. https://doi.org/10.1086/345321

Bardt, C. (2019). *Material and mind*. The MIT Press.

Barrett, L. (2019). William Grey Walter. In V. P. Glăveanu (Ed.), *The Palgrave encylopedia of the possible*. Palgrave MacMillan.

Bartholomew, T. T., & Brown, J. R. (2019). Entering the ethnographic mind: A grounded theory of using ethnography in psychological research. *Qualitative Research in Psychology*, 1–30. https://doi.org/10.1080/14780887.2019.1604927

Batchelder, W. H., & Alexander, G. E. (2012). Insight problem solving: A critical examination of the possibility of formal theory. *The Journal of Problem Solving*, *5*(1). https://doi.org/10.7771/1932-6246.1143

Bauer, M., & Gaskell, G. (2000). Towards public accountability: Beyond sampling, reliability and validity. In M. Bauer (Ed.), *Qualitative researching with text, image and sound* (pp. 337–350). SAGE Publications, Inc.

Baumeister, R. (2019). *Self-control, ego depletion, and social psychology's replication crisis* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/uf3cn

Beaty, R. E., Nusbaum, E. C., & Silvia, P. J. (2014). Does insight problem solving predict real-world creativity? *Psychology of Aesthetics, Creativity, and the Arts*, *8*(3), 287–292. https://doi.org/10.1037/a0035727

Beaty, R., & Johnson, D. R. (2020). *Automating creativity assessment with SemDis: An open platform for computing semantic distance* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/nwvps

Beghetto, R. A. (2020). Uncertainty. In V. P. Glăveanu (Ed.), *The Palgrave encylopedia of the possible*. Palgrave MacMillan.

Bennett, A., & Elman, C. (2006). Recent developments in case study methods. *Annual Review of Political Science*, *9*(1), 455–476. https://doi.org/10.1146/annurev.polisci.8.082103.104918

Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Blackwell.

Bickhard, M. H. (1992). Myths of science: Misconceptions of science in contemporary psychology. *Theory & Psychology*, *2*(3), 321–337. https://doi.org/10.1177/0959354392023004

Bikker, A. P., Atherton, H., Brant, H., Porqueddu, T., Campbell, J. L., Gibson, A., McKinstry, B., Salisbury, C., & Ziebland, S. (2017). Conducting a team-based multi-sited focused ethnography in primary care. *BMC Medical Research Methodology*, *17*(1). https://doi.org/10.1186/s12874-017-0422-5

Bilalić, M., Graf, M., Vaci, N., & Danek, A. H. (2019). The temporal dynamics of insight problem solving – restructuring might not always be sudden. *Thinking & Reasoning*, 1–37. https://doi.org/10.1080/13546783.2019.1705912

Bilda, Z., Gero, J. S., & Purcell, T. (2006). To sketch or not to sketch? That is the question. *Design Studies*, *27*(5), 587–613. https://doi.org/10.1016/j.destud.2006.02.002

Bingham, G. (1988). Task specific devices and the perceptual bottleneck. *Human Movement Science*, *7*, 225–264.

Björneborn, L. (2017). Three key affordances for serendipity: Toward a framework connecting environmental and personal factors in serendipitous encounters. *Journal of Documentation*, *73*, 1053–1081. https://doi.org/10.1108/JD-07-2016-0097

Boag, S. (2015). Personality assessment, 'construct validity', and the significance of theory. *Personality and Individual Differences*, *84*, 36–44. https://doi.org/10.1016/j.paid.2014.12.039

Bocanegra, B. R., Poletiek, F. H., Ftitache, B., & Clark, A. (2019a). Intelligent problem-solvers externalize cognitive operations. *Nature Human Behaviour*, *3*(2), 136–142. https://doi.org/10.1038/s41562-018-0509-y

Bocanegra, B. R., Poletiek, F. H., Ftitache, B., & Clark, A. (2019b). Intelligent problem-solvers externalize cognitive operations. *Nature Human Behaviour*, *3*(2), 136–142. https://doi.org/10.1038/s41562-018-0509-y

Bowden, E., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight. *Trends in Cognitive Sciences*, *9*(7), 322–328. https://doi.org/10.1016/j.tics.2005.05.012

Bowden, E. M., & Jung-Beeman, M. (2003). Aha! Insight experience correlates with solution activation in the right hemisphere. *Psychonomic Bulletin & Review*, *10*(3), 730–737. https://doi.org/10.3758/BF03196539

Breen, L., & Darlaston-Jones, D. (2010). Moving beyond the enduring dominance of positivism in psychological research: Implications for psychology in Australia. *Australian Psychologist*, *45*(1), 67–76. https://doi.org/10.1080/00050060903127481

Brenninkmeijer, J., Derksen, M., & Rietzschel, E. (2019). Informal laboratory practices in psychology. *Collabra: Psychology*, *5*(1), 45. https://doi.org/10.1525/collabra.221

Broadbent, D. (1958). *Perception and communication*. Pergamon Press.

Brown, J. R., & Fehige, Y. (2019). Thought Experiments. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Winter 2019 Edition)*. URL = <https://plato.stanford.edu/archives/win2019/entries/thought-experiment/>.

Bryant, A., & Charmaz, K. (Eds.). (2011). *The SAGE handbook of grounded theory*. SAGE.

Bryman, A. (1984). The debate about quantitative and qualitative research: A question of method or epistemology? *The British Journal of Sociology*, *1*, 75–92.

Bryman, A. (2007). Barriers to integrating quantitative and qualitative research. *Journal of Mixed Methods Research*, *1*(1), 8–22. https://doi.org/10.1177/2345678906290531

Butterworth, B. (1999). *What counts: How every brain is hardwired for math*. Free Press.

Campanario, J. M. (1996). Using Citation Classics to study the incidence of serendipity in scientific discovery. *Scientometrics*, *37*(1), 3–24. https://doi.org/10.1007/BF02093482

Campbell, D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review*, *67*, 380–400.

Campbell, J. I. D., & Clark, J. M. (1988). An encoding-complex view of cognitive number processing: Comment on McCloskey, Sokol, and Goodman (1986). *Journal of Experimental Psychology: General*, *117*(2), 204–214. https://doi.org/10.1037/0096-3445.117.2.204

Campbell, J. K., O'Rourke, M., & Slater, M. H. (Eds.). (2011). *Carving nature at its joints: Natural kinds in metaphysics and science*. MIT Press.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*(3), 404–431. https://doi.org/10.1037/0033-295X.97.3.404

Carter, J. A., & Peterson, M. (2017). The modal account of luck revisited. *Synthese*, *194*(6), 2175–2184. https://doi.org/10.1007/s11229-016-1047-7

Cary, M., & Carlson, R. A. (1999). External support and the development of problem-solving routines. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 1053–1070. https://doi.org/10.1037/0278-7393.25.4.1053

Charmaz, K. (2014). *Constructing grounded theory* (2nd edition). Sage.

Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology*, *15*(2), 181–195. https://doi.org/10.1207/S15326969ECO1502_5

Chemero, A. (2011). *Radical embodied cognitive science*. MIT Press.

Christensen, B. T., & Friis-Olivarius, M. (2020). How do initial ideas evolve into final ones? Exploring the cognitive size, structure and life of ideas using sticky notes. In B. T. Christensen, K. Halskov, & C. N. Klokmose (Eds.), *Sticky Creativity: Post-it note cognition, computers and design.* (pp. 53–75). Elsevier. https://doi.org/10.1016/B978-0-12-816566-9.00003-3

Chu, Y., & MacGregor, J. N. (2011). Human performance on insight problem solving: A review. *The Journal of Problem Solving*, *3*(2). https://doi.org/10.7771/1932-6246.1094

Chuderski, A., & Jastrzębski, J. (2018). Much ado about aha!: Insight problem solving is strongly related to working memory capacity and reasoning ability. *Journal of Experimental Psychology: General*, *147*(2), 257–281. https://doi.org/10.1037/xge0000378

Chuderski, A., Jastrzębski, J., Kroczek, B., Kucwaj, H., & Ociepka, M. (2020). Metacognitive experience on Raven's matrices versus insight problems. *Metacognition and Learning*. https://doi.org/10.1007/s11409-020-09236-6

Chuderski, A., Jastrzębski, J., & Kucwaj, H. (2020). How physical interaction with insight problems affects solution rates, hint use, and cognitive load. *British Journal of Psychology*. https://doi.org/10.1111/bjop.12442

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. MIT Press.

Clark, A. (2010). Material surrogacy and the supernatural: Reflections on the role of artefacts in 'offline'cognition. In L. Malafouris & C. Renfrew (Eds.), *The cognitive life of things: Recasting the boundaries of the mind* (pp. 23–28). McDonald Institute for Archaeological Research ; Oxbow Books.

Clark, A. (2014). *Mindware: An introduction to the philosophy of cognitive science*. Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19. https://doi.org/doi:10.1093/analys/58.1.7

Coffman, E. J. (2009). Does luck exclude control? *Australasian Journal of Philosophy*, *87*(3), 499–504. https://doi.org/10.1080/00048400802674677

Collins, H. M. (1975). The Seven Sexes: A study in the sociology of a phenomenon, or the replication of experiments in physics. *Sociology*, *9*(2), 205–224. https://doi.org/10.1177/003803857500900202

Colombo, M., Irvine, E., & Stapleton, M. (Eds.). (2019). *Andy Clark and his critics*. Oxford University Press.

Copeland, S. M. (2019). On serendipity in science: Discovery at the intersection of chance and wisdom. *Synthese*, *196*, 2385–2406. https://doi.org/10.1007/s11229-017-1544-3

Cornish, F., & Gillespie, A. (2009). A Pragmatist approach to the problem of knowledge in health psychology. *Journal of Health Psychology*, *14*(6), 800–809. https://doi.org/10.1177/1359105309338974

Craik, K. (1943). *The nature of explanation*. Cambridge University Press.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (Third Edition). SAGE.

Cummins, R. (2000). 'How does it work?' versus 'What are the laws?': Two conceptions of psychological explanation. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 118–144). MIT Press.

Cushen, P. J., & Wiley, J. (2012). Cues to solution, restructuring patterns, and reports of insight in creative problem solving. *Consciousness and Cognition*, *21*(3), 1166–1175. https://doi.org/10.1016/j.concog.2012.03.013

Danek, A. H. (2018). Magic tricks, sudden restructuring and the aha! experience. In F. Vallée-Tourangeau (Ed.), *Insight: On the origin of new ideas* (pp. 51–79). Routledge.

Danek, A. H., Fraps, T., von Maller, A., Grothe, B., & Ollinger, M. (2014). It's a kind of magic: What self-reports can reveal about the phenomenology of insight problem solving. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01408

Danek, A. H., Williams, J., & Wiley, J. (2020). Closing the gap: Connecting sudden representational change to the subjective Aha! experience in insightful problem solving. *Psychological Research*, *84*(1), 111–119. https://doi.org/10.1007/s00426-018-0977-8

De Jaegher, H. (2019). Loving and knowing: Reflections for an engaged epistemology. *Phenomenology and the Cognitive Sciences*. https://doi.org/10.1007/s11097-019-09634-5

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1–2), 1–42. https://doi.org/10.1016/0010-0277(92)90049-N

Dennett, D. C. (2014). *Intuition pumps and other tools for thinking*. Penguin Books.

Destefano, M., Lindstedt, J. K., & Gray, W. D. (2011). Use of complementary actions decreases with expertise. In L. Carlson, C. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.* (pp. 2709–2014). Cognitive Science Society.

Dixon, J. A., Kelty-Stephen, D. G., & Anastas, J. (2014). The embodied dynamics of problem solving. In L. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp. 160–170). Routledge.

Duncker, K. (1945). On problem-solving. (L. S. Lees, Trans.). *Psychological Monographs*, *58*(5), i–113. https://doi.org/10.1037/h0093599

Elder-Vass, D. (2012). *The reality of social construction*. Cambridge University Press.

Ellis, J. J., Glaholt, M. G., & Reingold, E. M. (2011). Eye movements reveal solution knowledge prior to insight. *Consciousness and Cognition*, *20*(3), 768–776. https://doi.org/10.1016/j.concog.2010.12.007

Engel, A. K., Maye, A., Kurthen, M., & König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends in Cognitive Sciences*, *17*(5), 202–209. https://doi.org/10.1016/j.tics.2013.03.006

Fayol, M., Barrouillet, P., & Marinthe, C. (1998). Predicting arithmetical achievement from neuro-psychological performance: A longitudinal study. *Cognition*, *68*(2), B63–B70. https://doi.org/10.1016/S0010-0277(98)00046-8

Fedor, A., Szathmáry, E., & Öllinger, M. (2015). Problem solving stages in the five square problem. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01050

Fernandez-Duque, D., & Johnson, M. L. (1999). Attention metaphors: How metaphors guide the cognitive psychology of attention. *Cognitive Science*, *23*(1), 83–116. https://doi.org/10.1207/s15516709cog2301_4

Fink, P. W., Foo, P. S., & Warren, W. H. (2009). Catching fly balls in virtual reality: A critical test of the outfielder problem. *Journal of Vision*, *9*(13), 14–14. https://doi.org/10.1167/9.13.14

Fioratou, E., & Cowley, S. J. (2009). Insightful thinking: Cognitive dynamics and material artifacts. *Pragmatics & Cognition*, *17*(3), 549–572. https://doi.org/10.1075/pc.17.3.04fio

Fischer, J.-P. (2010). Numerical performance increased by finger training: A fallacy due to regression toward the mean? *Cortex*, *46*(2), 272–273. https://doi.org/10.1016/j.cortex.2008.06.010

Fleck, J. I., & Weisberg, R. W. (2013). Insight versus analysis: Evidence for diverse methods in problem solving. *Journal of Cognitive Psychology*, *25*(4), 436–463. https://doi.org/10.1080/20445911.2013.779248

Fleming, M., & Maglio, P. (2015). How physical interaction helps performance in a Scrabble-like task. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 716–721). Cognitive Science Society.

Fodor, J. A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences*, *3*(1), 63–73. https://doi.org/10.1017/S0140525X00001771

Fodor, J. A. (1985). Fodor's guide to mental representation: The intelligent auntie's vade-mecum. *Mind*, *XCIV*(373), 76–100. https://doi.org/10.1093/mind/XCIV.373.76

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1–2), 3–71. https://doi.org/10.1016/0010-0277(88)90031-5

Foster, A. E., & Ellis, D. (2014). Serendipity and its study. *Journal of Documentation*, *70*, 1015–1038. https://doi.org/10.1108/JD-03-2014-0053

Friedlander, K. J., & Fine, P. A. (2016). The Grounded Expertise Components Approach in the novel area of cryptic crossword solving. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00567

Friedlander, K. J., & Fine, P. A. (2018). "The penny drops": Investigating insight through the medium of cryptic crosswords. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00904

Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind*. Oxford University Press.

Gavurin, E. (1967). Anagram solving and spatial aptitude. *Journal of Psychology*, *65*, 65–68.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Psychology Press.

Giere, R. (2006). Perspectival pluralism. In S. H. Kellert, H. E. Longino, & C. K. Waters (Eds.), *Scientific pluralism* (pp. 26–42). University of Minnesota Press.

Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, *1*(2), 198–218. https://doi.org/10.1177/2515245918771329

Gigerenzer, G. (2019). How to explain behavior? *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12480

Gigerenzer, G., & Goldstein, D. G. (1996). Mind as Computer: Birth of a Metaphor. *Creativity Research Journal*, *9*(2–3), 131–144. https://doi.org/10.1080/10400419.1996.9651168

Gilhooly, K. J. (2019). *Incubation in problem solving and creativity: Unconscious processes*. Routledge.

Gilhooly, K. J., Ball, L. J., & Macchi, L. (2015). Insight and creative thinking processes: Routine and special. *Thinking & Reasoning*, *21*(1), 1–4. https://doi.org/10.1080/13546783.2014.966758

Gilhooly, K. J., & Hay, D. (1977). Imagery, concreteness, age-of-acquisition, familiarity, and meaningfulness values for 205 five-letter words having single-solution anagrams. *Behavior Research Methods & Instrumentation*, *9*(1), 12–17. https://doi.org/10.3758/BF03202210

Gilhooly, K. J., & Johnson, C. E. (1978). Effects of solution word attributes on anagram difficulty: A regression analysis. *Quarterly Journal of Experimental Psychology*, *30*(1), 57–70. https://doi.org/10.1080/14640747808400654

Gilhooly, K. J., & Murphy, P. (2005). Differentiating insight from non-insight problems. *Thinking & Reasoning*, *11*(3), 279–302. https://doi.org/10.1080/13546780442000187

Gilhooly, K. J., & Webb, M. E. (2018). Working memory in insight problem solving. In F. Vallée-Tourangeau (Ed.), *Insight: On the origins of new ideas.* (pp. 105–120). Routledge.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for Qualitative Research*. Aldine Publishing Company.

Glăveanu, V. P. (2012). What can be done with an egg? Creativity, material objects, and the theory of affordances. *The Journal of Creative Behavior*, *46*, 192–208. https://doi.org/10.1002/jocb.13

Glăveanu, V. P. (2013). Creativity and folk art: A study of creative action in traditional craft. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(2), 140–154.

Glăveanu, V. P. (2020). A sociocultural theory of creativity: Bridging the social, the material, and the psychological. *Review of General Psychology*, 108926802096176. https://doi.org/10.1177/1089268020961763

Glăveanu, V. P., & Beghetto, R. A. (2020). Creative experience: A non-standard definition of creativity. *Creativity Research Journal*, 1–6. https://doi.org/10.1080/10400419.2020.1827606

Glaveanu, V. P., Hanchett Hanson, M., Baer, J., Barbot, B., Clapp, E. P., Corazza, G. E., Hennessey, B., Kaufman, J. C., Lebuda, I., Lubart, T., Montuori, A., Ness, I. J., Plucker, J., Reiter-Palmon, R., Sierra, Z., Simonton, D. K., Neves-Pereira, M. S., & Sternberg, R. J. (2020). Advancing creativity theory and research: A socio-cultural manifesto. *The Journal of Creative Behavior*, *54*(3), 741–745. https://doi.org/10.1002/jocb.395

Glăveanu, V. P., Lubart, T., Bonnardel, N., Botella, M., Biaisi, P.-M. de, Desainte-Catherine, M., Georgsdottir, A., Guillou, K., Kurtag, G., Mouchiroud, C., Storme, M., Wojtczuk, A., & Zenasni, F. (2013). Creativity as action: Findings from five creative domains. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00176

Goel, A. (2019). A cognitive reformation. *Topics in Cognitive Science*, *11*(4), 892–901. https://doi.org/10.1111/tops.12469

Goel, V. (2010). Neural basis of thinking: Laboratory problems versus real-world problems. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(4), 613–621. https://doi.org/10.1002/wcs.71

Goldinger, S. D., Papesh, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The poverty of embodied cognition. *Psychonomic Bulletin & Review*, *23*(4), 959–978. https://doi.org/10.3758/s13423-015-0860-1

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review*, *100*(2), 279–297. https://doi.org/10.1037/0033-295X.100.2.279

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, *5*(6), 664–674. https://doi.org/10.1177/1745691610388764

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science*, *20*(3), 267–272. https://doi.org/10.1111/j.1467-9280.2009.02297.x

Gozli, D. G. (2017). Behaviour versus performance: The veiled commitment of experimental psychology. *Theory & Psychology*, *27*(6), 741–758. https://doi.org/10.1177/0959354317728130

Gracia-Bafalluy, M., & Noel, M. (2008). Does finger training increase young children's numerical performance? *Cortex*, *44*(4), 368–375. https://doi.org/10.1016/j.cortex.2007.08.020

Griffith, M. (2010). Why agent caused actions are not lucky. *North American Philosophical Publications*, *47*(1), 43–56.

Guildford, J. P., & Hoepfner, R. (1971). *The analysis of intelligence*. McGraw-Hill.

Guthrie, L. G., & Vallée-Tourangeau, F. (2015). Interactivity and mental arithmetic: Coupling mind and world transforms and enhances performance. *Studies in Logic, Grammar and Rhetoric*, *41*(1), 41–59. https://doi.org/10.1515/slgr-2015-0019

Hacking, I. (1982). Experimentation and scientific realism. *Philsophical Topics*, *13*(1), 71–87.

Heath, C., Hindmarsh, J., & Luff, P. (2010). *Video in qualitative research: Analysing social interaction in everyday life*. SAGE.

Henok, N., Vallée-Tourangeau, F., & Vallée-Tourangeau, G. (2020). Incubation and interactivity in insight problem solving. *Psychological Research*, *84*(1), 128–139. https://doi.org/10.1007/s00426-018-0992-9

Hoad, T. F. (Ed.). (1993). *The Concise Oxford dictionary of English etymology*. Oxford University Press.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on*

Computer-Human Interaction (TOCHI), *7*(2), 174–196. https://doi.org/10.1145/353485.353487

Hollnagel, E. (2007). The elusiveness of cognition. In B. Wallace, A. Ross, J. Davies, & T. Anderson (Eds.), *The mind, the body, and the world: Psychology after cognitivism?* Imprint Academic.

Holyoak, K. J., & Morrison, R. G. (Eds.). (2013). *The Oxford handbook of thinking and reasoning*. Oxford University Press.

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, *137*(2), 297–315. https://doi.org/10.1037/a0022128

Hostetter, A. B., & Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*, *26*(3), 721–752. https://doi.org/10.3758/s13423-018-1548-0

Hunter, I. (1959). The solving of five letter anagrams. *British Journal of Psychology*, *50*(3), 193–206.

Hurley, S. L. (2001). Perception and action: Alternative views. *Synthese*, *129*(1), 3–40. https://doi.org/10.1023/A:1012643006930

Hurley, S. L. (2002). *Consciousness in action*. Harvard University Press.

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Hutchins, E. (2010a). Enaction, imagination and insight. In J. Stewart, O. Gapenne, & E. Di Paolo (Eds.), *Enaction: Towards a new paradigm for cognitive science* (pp. 424–450). MIT Press.

Hutchins, E. (2010b). Cognitive ecology. *Topics in Cognitive Science*, *2*(4), 705–715. https://doi.org/10.1111/j.1756-8765.2010.01089.x

Ingold, T. (2010). The textility of making. *Cambridge Journal of Economics*, *34*(1), 91–102. https://doi.org/10.1093/cje/bep042

James, W. (1890). *The principles of psychology*. retrieved from http://library.manipaldubai.com/DL/the_principles_of_psychology_vol_I.pdf.

JASP Team. (2020). *JASP* (0.14) [Computer software].

Jones, G. (2003). Testing two cognitive theories of insight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 1017–1027. https://doi.org/10.1037/0278-7393.29.5.1017

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences*, *4*(1), 39–103. https://doi.org/10.1207/s15327809jls0401_2

Kahneman, D. (2012). *Thinking, fast and slow*. Penguin Books.

Kaplan, C., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology*, *22*, 374–419.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. MIT Press.

Karwowski, M., & Lebuda, I. (2016). The big five, the huge two, and creative self-beliefs: A meta-analysis. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(2), 214–232. https://doi.org/10.1037/aca0000035

Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four C model of creativity. *Review of General Psychology*, *13*, 1–12. https://doi.org/10.1037/a0013688

Kaufmann, G. (2003). What to measure? A new look at the concept of creativity. *Scandinavian Journal of Research*, *47*(3), 235–251. https://doi.org/10.1080/00313830320000079227

Kelle, U. (2011). The development of categories: Different approaches to Grounded Theory. In A. Bryant & K. Charmaz (Eds.), *The SAGE handbook of grounded theory* (pp. 191–213). SAGE.

Ketokivi, M., & Choi, T. (2014). Renaissance of case research as a scientific method. *Journal of Operations Management*, *32*(5), 232–240. https://doi.org/10.1016/j.jom.2014.03.004

Kirchhoff, M. D. (2012). Extended cognition and fixed properties: Steps to a third-wave version of extended cognition. *Phenomenology and the Cognitive Sciences*, *11*(2), 287–308. https://doi.org/10.1007/s11097-011-9237-8

Kirchhoff, M. D. (2018). *Extended consciousness and predictive processing: A third wave view*. Routledge.

Kirsh, D. (2009). Problem solving and situated cognition. In P. Robbins & M. Ayede (Eds.), *The Cambridge handbook of situated cognition* (pp. 264–306). Cambridge University Press.

Kirsh, D. (2014). The importance of chance and interactivity in creativity. *Pragmatics & Cognition*, *22*(1), 5–26. https://doi.org/10.1075/pc.22.1.01kir

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18*(4), 513–549. https://doi.org/10.1207/s15516709cog1804_1

Kiverstein, J. (2018). Extended cognition. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition* (pp. 18–40). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.013.2

Knight, D., & Muncer, S. J. (2011). Type and token bigram frequencies for two-through nine-letter words and the prediction of anagram difficulty. *Behavior Research Methods*, *43*, 491–498. https://doi.org/10.3758/s13428-011-0068-x

Knoblauch, H., Soeffner, H.-G., Raab, J., & Schnettler, B. (2013). *Video analysis: Methodology and methods. Qualitative audiovisual data analysis in sociology*. Peter Lang GmbH, Internationaler Verlag der Wissenschaften.

Knorr Cetina, K. (1992). The couch, the cathedral and the laboratory: On the relationship between the experiment and the laboratory in science. In A. Pickering (Ed.), *Sceince as practice and culture* (pp. 113–138). University of Chicago Press.

Koriat, A., & Goldsmith, M. (1996). Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioral and Brain Sciences*, *19*(2), 167–188. https://doi.org/10.1017/S0140525X00042114

Kounios, J., Fleck, J. I., Green, D. L., Payne, L., Stevenson, J. L., Bowden, E. M., & Jung-Beeman, M. (2009). The origins of insight in resting-state brain activity. *Neuropsychologia*, *46*(1), 281–291. https://doi.org/10.1016/j.neuropsychologia.2007.07.013

Krug, M. (2018). *Collecting audio-visual data of theatre rehearsals. (Non-)intrusive practices of preparing mobile eye-tracking glasses during ongoing workplace interactions*. https://doi.org/10.13140/RG.2.2.21186.89281

Kuiken, D., Wild, T. C., & Schopflocher, D. (1992). Positivist conceptions of induction and the rejection of classificatory methods in psychological research. In C. W. Tolman (Ed.), *Positivism in Psychology* (pp. 47–56). Springer New York. https://doi.org/10.1007/978-1-4612-4402-8_5

Lahlou, S. (2011). How can we capture the subject's perspective? An evidence-based approach for the social scientist. *Social Science Information*, *50*(3–4), 607–655. https://doi.org/10.1177/0539018411411033

Lakoff, G., & Johnson, M. (2010). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. Basic Books.

Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Harvard University Press.

Laurence, P. G., Mecca, T. P., Serpa, A., Martin, R., & Macedo, E. C. (2018). Eye movements and cognitive strategy in a fluid intelligence test: Item type analysis. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00380

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge University Press. https://doi.org/10.1017/CBO9780511609268

Lee, K., & Ashton, M. C. (2018). Psychometric properties of the HEXACO-100. *Assessment*, *25*, 543–556.

Leonelli, S. (2018). Rethinking reproducibility as a criterion for research quality. In L. Fiorito, S. Scheall, & C. E. Suprinyak (Eds.), *Research in the History of Economic Thought and Methodology* (Vol. 36, pp. 129–146). Emerald Publishing Limited. https://doi.org/10.1108/S0743-41542018000036B009

Long, H. (2014). An empirical review of research methodologies and methods in creativity studies (2003–2012). *Creativity Research Journal*, *26*(4), 427–438. https://doi.org/10.1080/10400419.2014.961781

Loughlin, V. (2020). The Extended Mind. In V. P. Glăveanu (Ed.), *The Palgrave encylopedia of the possible*.

Love, N. (2004). Cognition and the language myth. *Language Sciences*, *26*(6), 525–544. https://doi.org/10.1016/j.langsci.2004.09.003

Lycan, W. G., & Prinz, J. J. (2008). Ontology: The indentity theory and functionlism. Introduction. In W. G. Lycan & J. J. Prinz (Eds.), *Mind and cognition: An anthology* (pp. 3–15). Blackwell Pub. Ltd.

Mace, M.-A., & Ward, T. (2002). Modeling the creative Process: A grounded theory analysis of creativity in the domain of art making. *Creativity Research Journal*, *14*(2), 179–192. https://doi.org/10.1207/S15326934CRJ1402_5

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2001). Information processing and insight: A process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 176–201. https://doi.org/10.1037//0278-7393.27.1.176

Maglio, P. P., Matlock, T., Raphaely, D., Chernicky, B., & Kirsh, D. (1999). Interactive skill in Scrabble. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the 21st Annual Conference of the Cognitive Science Society* (pp. 326–330). Lawrence Erlbaum Associates, Publishers.

Mahon, B. Z. (2015). What is embodied about cognition? *Language, Cognition and Neuroscience*, *30*(4), 420–429. https://doi.org/10.1080/23273798.2014.987791

Makri, S., & Blandford, A. (2012). Coming across information serendipitously – Part 1: A process model. *Journal of Documentation*, *68*, 684–705. https://doi.org/10.1108/00220411211256030

Malafouris, L. (2008a). At the potter's wheel: An argument for material agency. In C. Knappett & L. Malafouris (Eds.), *Material Agency* (pp. 19–36). Springer US. https://doi.org/10.1007/978-0-387-74711-8_2

Malafouris, L. (2008b). Beads for a plastic mnd: The 'Blind Man's Stick' (BMS) hypothesis and the active nature of material culture. *Cambridge Archaeological Journal*, *18*(3), 401–414. https://doi.org/10.1017/S0959774308000449

Malafouris, L. (2010). The brain–artefact interface (BAI): A challenge for archaeology and cultural neuroscience. *Social Cognitive and Affective Neuroscience*, *5*(2–3), 264–273. https://doi.org/10.1093/scan/nsp057

Malafouris, L. (2013). *How things shape the mind: A theory of material engagement*. MIT Press.

Malafouris, L. (2014). Creative thinging: The feeling of and for clay. *Pragmatics & Cognition*, *22*(1), 140–158. https://doi.org/10.1075/pc.22.1.08mal

Malafouris, L. (2015). Metaplasticity and the primacy of material engagement. *Time and Mind*, *8*(4), 351–371. https://doi.org/10.1080/1751696X.2015.1111564

Malafouris, L. (2018). Bringing things to mind: 4Es and Material Engagement. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition* (pp. 754–772). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.013.40

Malafouris, L. (2020). Thinking as "Thinging": Psychology with things. *Current Directions in Psychological Science*, *29*(1), 3–8. https://doi.org/10.1177/0963721419873349

Mandler, G. (2011). *A history of modern experimental psychology: From James and Wundt to cognitive science*. MIT CogNet.

March, P. L. (2019). Playing with clay and the uncertainty of agency. A Material Engagement Theory perspective. *Phenomenology and the Cognitive Sciences*, *18*(1), 133–151. https://doi.org/10.1007/s11097-017-9552-9

March, P. L., & Glăveanu, V. P. (2020). Craft. In *Encyclopedia of Creativity* (pp. 215–221). Elsevier. https://doi.org/10.1016/B978-0-12-809324-5.23718-8

March, P. L., Ross, W., & Vallée-Tourangeau, F. (in preparation). Cognition as choreography. The relationship between eye movements and the morphology of a jomon flame-style pot. *Ethnoarcheology*.

Martin, K., & Quan-Haase, A. (2016). The role of agency in historians' experiences of serendipity in physical and digital information environments. *Journal of Documentation*, *72*, 1008–1026. https://doi.org/10.1108/JD-11-2015-0144

Martin, L., & Schwartz, D. L. (2009). Prospective Adaptation in the Use of External Representations. *Cognition and Instruction*, *27*(4), 370–400.

Martin, P., & Bateson, P. (1993). *Measuring behaviour: An introductory guide*. Cambridge University Press.

Max Planck Institute for Psycholinguistics. (2018). *ELAN* (5.4) [Computer software]. The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Max Planck Institute for Psycholinguistics. (2019). *ELAN* (5.5) [Computer software]. The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

May, A. (2011). Experience-dependent structural plasticity in the adult human brain. *Trends in Cognitive Sciences*, *15*(10), 475–482. https://doi.org/10.1016/j.tics.2011.08.002

McCay-Peet, L., Toms, E. G., & Kelloway, E. K. (2015). Examination of relationships among serendipity, the environment, and individual differences. *Information Processing & Management*, *51*(4), 391–412. https://doi.org/10.1016/j.ipm.2015.02.004

McClelland, J. L., & Rumelhart, D. E. (1987). *Parallel distributed processing. Explorations in the microstructure of cognition: Psychological and biological models. Vol 2.* MIT Press.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, *44*(1–2), 107–157. https://doi.org/10.1016/0010-0277(92)90052-J

McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition*, *4*(2), 171–196. https://doi.org/10.1016/0278-2626(85)90069-7

Menary, R. (Ed.). (2010). *The extended mind*. MIT Press.

Merleau-Ponty, M. (1945). *Phenomenology of perception*. Routledge and Kegan Paul (trans C. Smith).

Metcalfe, J. (1986). Premonitions of insight predict impending error. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(4), 623–634. https://doi.org/10.1037/0278-7393.12.4.623

Moeller, K., Martignon, L., Wessolowski, S., Engel, J., & Nuerk, H.-C. (2011). Effects of finger counting on numerical development ? The opposing views of neurocognition and mathematics education. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00328

Montuori, A., & Purser, R. E. (1995). Deconstructing the lone genius myth: Toward a contextual view of creativity. *Journal of Humanistic Psychology*, *35*, 69–112.

Neth, H., & Payne, S. J. (2011). Interactive coin addition: How hands can help us think. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proceedings of the thirty-third annual conference of the cognitive science society* (pp. 279–284). Cognitive Science Society.

Neves-Pereira, M. S. (2019). Microgenetic analysis and creativity: Analyzing psychological change processes. In I. Lebuda & V. P. Glăveanu (Eds.), *The Palgrave Handbook of Social Creativity Research* (pp. 145–161). Springer International Publishing. https://doi.org/10.1007/978-3-319-95498-1_10

Newen, A., Gallagher, S., & De Bruin, L. (2018). 4E cognition: Historical roots, key concepts, and central Issues. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The*

Oxford handbook of 4E cognition (pp. 2–16). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.013.1

Niglas, K. (2010). The multidimensional model of research methodology: An integrated set of continua. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social & behavioral Research* (pp. 215–236). SAGE. https://doi.org/10.4135/9781506335193.n9

Noë, A. (2006). *Action in perception*. MIT Press.

Noë, A. (2010). *Out of our heads: Why you are not your brain, and other lessons from the biology of consciousness*. Hill and Wang.

Noë, A. (2012). *Varieties of presence*. Harvard University Press.

Novick, L. R., & Sherman, S. J. (2003). On the nature of insight solutions: Evidence from skill differences in anagram solution. *The Quarterly Journal of Experimental Psychology*, *56*(2), 351–382. https://doi.org/10.1080/02724980244000288

Novick, L. R., & Sherman, S. J. (2008). The effects of superficial and structural information on online problem solving for good versus poor anagram solvers. *Quarterly Journal of Experimental Psychology*, *61*(7), 1098–1120. https://doi.org/10.1080/17470210701449936

Nunes, T., Schliemann, A. D., & Carraher, D. W. (1993). *Street mathematics and school mathematics*. Cambridge University Press.

Núñez, R., Allen, M., Gao, R., Miller Rigoli, C., Relaford-Doyle, J., & Semenuks, A. (2019). What happened to cognitive science? *Nature Human Behaviour*, *3*(8), 782–791. https://doi.org/10.1038/s41562-019-0626-2

Ohlsson, S. (1984a). Restructuring revisited: I: Summary and critique of the Gestalt theory of problem solving. *Scandinavian Journal of Psychology*, *25*, 65–78.

Ohlsson, S. (1984b). Restructuring revisited: II: An information processing theory of restructuring. *Scandinavian Journal of Psychology*, *25*, 117–129.

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. T. Keane & K. J. Gilhooly (Eds.), *Advances in the psychology of thinking* (pp. 1–44). Harvester Wheatsheaf.

Ohlsson, S. (2011). *Deep learning: How the mind overrides experience*. Cambridge University Press.

Ohlsson, S. (2018). The dialectic between routine and creative cognition. In F. Vallée-Tourangeau (Ed.), *Insight: On the origins of new ideas* (pp. 8–28).

Onwuegbuzie, A., & Combs, J. (2011). Data analysis in mixed research: A primer. *International Journal of Education*, *3*(1), Article e13. https://doi.org/doi:10.5296/ije.v3i1.618

Onwuegbuzie, A. J., & Leech, N. L. (2005). On becoming a pragmatic researcher: The Importance of combining quantitative and qualitative research methodologies. *International Journal of Social Research Methodology*, *8*(5), 375–387. https://doi.org/10.1080/13645570500402447

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization Studies*, *28*(9), 1435–1448. https://doi.org/10.1177/0170840607081138

Ormerod, T. C., MacGregor, J. N., & Chronicle, E. P. (2002). Dynamics and constraints in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 791–799. https://doi.org/10.1037//0278-7393.28.4.791

Ormerod, T. C., MacGregor, J. N., Chronicle, E. P., Dewald, A. D., & Chu, Y. (2013). Act first, think later: The presence and absence of inferential planning in problem solving. *Memory & Cognition*, *41*(7), 1096–1108. https://doi.org/10.3758/s13421-013-0318-5

Overmann, K. A. (2018). Constructing a concept of number. *Journal of Numerical Cognition*, *4*(2), 464–493. https://doi.org/10.5964/jnc.v4i2.161

Pallasmaa, J. (2012). *The eyes of the skin: Architecture and the senses* (3. ed). Wiley.

Peirce, C. S. (1868). Some consequences of four incapacities. *The Journal of Speculative Philosophy*, *2*(3), 140–157.

Peirce, C. S. (1878). How to make our ideas clear. *Popular Science Monthly*, *12*, 286–302.

Pernu, T. K. (2017). The five marks of the mental. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01084

Peterson, D. (2016). The baby factory: Difficult research objects, disciplinary standards, and the production of statistical significance. *Socius: Sociological Research for a Dynamic World*, *2*, 237802311562507. https://doi.org/10.1177/2378023115625071

Piccinini, G. (2012). Computationalism. In E. Margolis, R. Samuels, & S. P. Stich (Eds.), *The Oxford handbook of philosophy of cognitive science* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195309799.013.0010

Pickering, Andrew. (1995). *The mangle of practice: Time, agency, and science*. University of Chicago Press.

Pilgrim, D. (2020). *Critical realism for psychologists*.

Pritchard, D. (2014). The modal account of luck. *Metaphilosophy*, *45*(4–5), 594–619. https://doi.org/10.1111/meta.12103

Radder, H. (Ed.). (2003). *The philosophy of scientific experimentation*. University of Pittsburgh Press.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, *20*(2), 110–122. https://doi.org/10.1016/j.lindif.2009.10.005

Ratcliffe, M. (2018). Perception, exploration, and the primacy of touch. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford handbook of 4E cognition* (pp. 280–300). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.013.14

Reed, S. K. (2016). The structure of ill-structured (and well-structured) problems revisited. *Educational Psychology Review*, *28*(4), 691–716. https://doi.org/10.1007/s10648-015-9343-1

Reiter-Palmon, R., Forthmann, B., & Barbot, B. (2019). Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 144–152. https://doi.org/10.1037/aca0000227

Reiter-Palmon, R., & Robinson, E. J. (2009). Problem identification and construction: What do we know, what is the future? *Psychology of Aesthetics, Creativity, and the Arts*, *3*(1), 43–47. https://doi.org/10.1037/a0014629

Renfrew, C. (1993). Cognitive archaeology: Some thoughts on the archaeology of thought. *Cambridge Archaeological Journal*, *3*(2), 248–250.

Rescher, N. (1995). *Luck: The brilliant randomness of everyday life*. Farrar, Straus and Giroux.

Rietveld, E., Denys, D., & Van Westen, M. (2018). Ecological-enactive cognition as engaging with a field of relevant affordances: The Skilled Intentionality Framework (SIF). In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford handbook of 4E cognition* (pp. 40–70). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.013.3

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Robbins, P. (2017). Modularity of mind. *The Stanford Encyclopedia of Philosophy (Winter 2017 Edition)*. <https://plato.stanford.edu/archives/win2017/entries/modularity-mind/>.

Robinson, O. (2011). The idiographic/nomothetic dichotomy: Tracing historical origins of contemporary confusions. *History & Philosophy of Psychology*, *13*(2), 32–39.

Rocca, E., & Anjum, R. L. (2020). Complexity, reductionism and the biomedical model. In R. L. Anjum, S. Copeland, & E. Rocca (Eds.), *Rethinking causality, complexity and evidence for the unique patient a CauseHealth resource for healthcare professionals and the clinical encounter*. Springer Open.

Rorty, R. (1991). *Objectivity, relativism, and truth*. Cambridge University Press.

Ross, W. (forthcominga). Microserendipity and creative processes. In W. Ross & S. Copeland (Eds.), *The Art of Serendipity*. Palgrave MacMillan.

Ross, W. (forthcomingb). Serendipity and Cognitive Science. In S. M. Copeland, M. Sand, & W. Ross (Eds.), *Serendipity Science*. Springer.

Ross, W. (2020). Serendipity. In V. P. Glăveanu (Ed.), *The Palgrave encylopedia of the possible*. Palgrave MacMillan.

Ross, W., & Arfini, S. (Forthcoming). Serendipity and creative cognition. In L. J. Ball & F. Vallée-Tourangeau (Eds.), *Routledge handbook of creative cognition*.

Ross, W., & Vallée-Tourangeau, F. (2018). For a good poet's made, as well as born: The relational ontology of Shakespeare's genius. *Creativity. Theories – Research - Applications*, *5*(1), 26–40. https://doi.org/10.1515/ctra-2018-0002

Ross, W., & Vallée-Tourangeau, F. (2020). Catch that word: Interactivity, serendipity and verbal fluency in a word production task. *Psychological Research*. https://doi.org/10.1007/s00426-019-01279-y

Ross, W., Vallée-Tourangeau, F., & Glăveanu, V. P. (2020). Collaboration. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of creativity*.

Ross, W., Vallée-Tourangeau, F., & Van Herwegen, J. (2020). Mental arithmetic and interactivity: The effect of manipulating external number representations on older children's mental arithmetic success. *International Journal of Science and Mathematics Education*, 985–1000. https://doi.org/10.1007/s10763-019-09978-z

Rossman, G. B., & Wilson, B. L. (1985). Numbers and words: Combining quantitative and qualitative methods in a single large-scale evaluation study. *Evaluation Review*, *9*(5), 627–643. https://doi.org/10.1177/0193841X8500900505

Rowlands, M. (2010). *The new science of the mind: From extended mind to embodied phenomenology*. MIT Press.

Royston, R., & Reiter-Palmon, R. (2019). Creative self-efficacy as mediator between creative mindsets and creative problem-solving. *The Journal of Creative Behavior*, *53*(4), 472–481. https://doi.org/10.1002/jocb.226

Rubin, V. L., Burkell, J., & Quan-Haase, A. (2011). Facets of serendipity in everyday chance encounters: A grounded theory approach to blog analysis. *Information Research*, *16*. https://doi.org/[Available at http://InformationR.net/ir/16-3/paper488.html]

Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal*, *24*(1), 66–75. https://doi.org/10.1080/10400419.2012.652929

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, *24*, 92–96. https://doi.org/10.1080/10400419.2012.650092

Rupert, R. D. (2004). Challenges to the hypothesis of extended cognition. *The Journal of Philsophy*, *101*(8), 389–428. https://doi.org/10.5840/jphil2004101826

Rupert, R. D. (2009). *Cognitive systems and the extended mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195379457.001.0001

Ruthven, K., & Chaplin, D. (1997). The calculator as cognitive tool: Upper primary pupils tackling a realistic number problem. *International Journal of Computers for Mathematical Learning*, *2*(2), 93–124. https://doi.org/10.1023/A:1009747024970

Saldaña, J. (2016). *The coding manual for qualitative researchers*. SAGE.

Sand, M. (2020). Did Alexander Fleming deserve the Nobel prize? *Science and Engineering Ethics*, *26*(2), 899–919. https://doi.org/10.1007/s11948-019-00149-5

Scheiter, K., & Gerjets, P. (2003). Sequence effects in solving knowledge-rich problems: The ambiguous role of surface similarities. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 1035–1040). Cognitive Science Society.

Seawright, J., & Gerring, J. (2008). Case selection techniques in case study research: A menu of qualitative and quantitative options. *Political Research Quarterly*, *61*(2), 294–308. https://doi.org/10.1177/1065912907313077

Seifert, C. M., Meyer, D. E., Davidson, N., Patalano, A. L., & Yaniv, I. (1994). Demystification of cognitive insight: Opportunistic assimilation and the prepared-mind hypothesis. In R. J. Sternberg & J. Davidson (Eds.), *The nature of insight* (pp. 65–124). MIT Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simon, H. A. (1965). *The sciences of the artificial* (3. ed.,). MIT Press.

Simon, H. A. (1977). The structure of ill-structured problems. In R. S. Cohen & M. W. Wartofsky (Eds.), *Models of discovery* (Vol. 54, pp. 304–325). Springer Netherlands. https://doi.org/10.1007/978-94-010-9521-1_17

Simonton, D. K. (2010). Creative thought as blind-variation and selective-retention: Combinatorial models of exceptional creativity. *Physics of Life Reviews*, *7*(2), 156–179. https://doi.org/10.1016/j.plrev.2010.02.002

Sjöholm, J. (2014). The art studio as archive: Tracing the geography of artistic potentiality, progress and production. *Cultural Geographies*, *21*, 505–514. https://doi.org/10.1177/1474474012473060

Smith, J. A., Harré, R., & Langenhove, L. van. (1995). Idiography and the case study. In J. A. Smith, R. Harré, & L. van Langenhove (Eds.), *Rethinking psychology* (pp. 59–70). Sage Publications.

Smythe, W. E. (1992). Positivism and the prospects for cognitive science. In C. W. Tolman (Ed.), *Positivism in psychology* (pp. 103–118). Springer New York. https://doi.org/10.1007/978-1-4612-4402-8_9

Steffensen, S. V. (2016). Cognitive probatonics: Towards an ecological psychology of cognitive particulars. *New Ideas in Psychology*, *42*, 29–38. https://doi.org/10.1016/j.newideapsych.2015.07.003

Steffensen, S. V. (2017). Human interactivity: Problem-solving, solution-probing, and verbal patterns in the wild. In S.J Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain* (pp. 85–113). Springer. https://doi.org/10.1007/978-3-319-49115-8

Steffensen, S. V., & Vallée-Tourangeau, F. (2018). An ecological perspective on insight problem solving. In F. Vallée-Tourangeau (Ed.), *Insight: On the origins of new ideas.* (pp. 169–190). Routledge.

Steffensen, S. V., Vallée-Tourangeau, F., & Vallée-Tourangeau, G. (2016). Cognitive events in a problem-solving task: A qualitative method for investigating interactivity in the 17 Animals problem. *Journal of Cognitive Psychology*, *28*(1), 79–105. https://doi.org/10.1080/20445911.2015.1095193

Stuart, M. T. (2020). Thought experiments. In V. P. Glăveanu (Ed.), *The Palgrave encylopedia of the possible*.

Sugarman, J. (2017). Psychologism as a style of reasoning and the study of persons. *New Ideas in Psychology*, *44*, 21–27. https://doi.org/10.1016/j.newideapsych.2016.11.008

Sutton, J. (2006). Distributed cognition: Domains and dimensions. *Pragmatics & Cognition*, *14*(2), 235–247. https://doi.org/10.1075/pc.14.2.05sut

Sutton, J. (2010). Exograms and interdisciplinarity: History, the extended mind, and the civilising process. In R. Menary (Ed.), *The extended mind* (pp. 189–227). MIT Press.

Tanggaard, L., & Beghetto, R. A. (2015). Ideational pathways: Toward a new approach for studying the life of ideas. *Creativity. Theories – Research - Applications*, *2*(2), 129–144. https://doi.org/10.1515/ctra-2015-0017

Thagard, P. (2005). *Mind: Introduction to cognitive science*. MIT Press.

Thagard, P. (2014). Thought experiments considered harmful. *Perspectives on Science*, *22*(2), 288–305. https://doi.org/10.1162/POSC_a_00131

Thompson, E. (2016). Introduction to the revised edition. In F. J. Varela, E. Thompson, & E. Rosch, *The embodied mind: Cognitive science and human experience* (revised edition, pp. xvii–xxxiii). MIT Press.

Tinio, P. P. L. (2013). From artistic creation to aesthetic reception: The mirror model of art. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(3), 265–275. https://doi.org/10.1037/a0030872

Toomela, A. (2020). Psychology today: Still in denial, still outdated. *Integrative Psychological and Behavioral Science*, *54*(3), 563–571. https://doi.org/10.1007/s12124-020-09534-3

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*(4), 242–248. https://doi.org/10.1080/17470216008416732

Tuffiash, M., Roring, R. W., & Ericsson, K. A. (2007). Expert performance in SCRABBLE: Implications for the study of the structure and acquisition of complex skills. *Journal of Experimental Psychology: Applied*, *13*, 124–134. https://doi.org/10.1037/1076-898X.13.3.124

Urquhart, C., Lehmann, H., & Myers, M. D. (2009). Putting the ‘theory’ back into grounded theory: Guidelines for grounded theory studies in information systems: Guidelines for grounded theory studies in information systems. *Information Systems Journal*, *20*(4), 357–381. https://doi.org/10.1111/j.1365-2575.2009.00328.x

Vallée-Tourangeau, F. (2013). Interactivity, efficiency, and individual differences in mental arithmetic. *Experimental Psychology*, *60*, 302–311. https://doi.org/10.1027/1618-3169/a000200

Vallée-Tourangeau, F. (2014). Insight, interactivity and materiality. *Pragmatics & Cognition*, *22*(1), 27–44. https://doi.org/10.1075/pc.22.1.02val

Vallée-Tourangeau, F., Euden, G., & Hearn, V. (2011). Einstellung defused: Interactivity and mental set. *Quarterly Journal of Experimental Psychology*, *64*(10), 1889–1895. https://doi.org/10.1080/17470218.2011.605151

Vallée-Tourangeau, F., & March, P. L. (2019). Insight out: Making creativity visible. *The Journal of Creative Behavior*. https://doi.org/10.1002/jocb.409

Vallée-Tourangeau, F., Ross, W., Ruffatto Rech, R., & Vallée-Tourangeau, G. (2020). Insight as discovery. *Journal of Cognitive Psychology*, 1–20. https://doi.org/10.1080/20445911.2020.1822367

Vallée-Tourangeau, F., Sirota, M., & Vallée-Tourangeau, G. (2016). Interactivity mitigates the impact of working memory depletion on mental arithmetic performance. *Cognitive Research: Principles and Implications*, 1–26. https://doi.org/10.1186/s41235-016-0027-2

Vallée-Tourangeau, F., Steffensen, S. V., Vallée-Tourangeau, G., & Sirota, M. (2016). Insight with hands and things. *Acta Psychologica*, *170*, 195–205. https://doi.org/doi: 10.1016/j.actpsy.2016.08.006

Vallée-Tourangeau, F., & Vallée-Tourangeau, G. (2014). Diagrams, jars, and matchsticks: A systemicist's toolkit. *Pragmatics & Cognition*, *22*(2), 187–205. https://doi.org/10.1075/pc.22.2.02val

Vallée-Tourangeau, F., & Vallée-Tourangeau, G. (2020). Mapping systemic resources in problem solving. *New Ideas in Psychology*. https://doi.org/10.1016/j.newideapsych.2020.100812

Vallée-Tourangeau, F., & Wrightman, M. (2010). Interactive skills and individual differences in a word production task. *AI & SOCIETY*, *25*, 433–439. https://doi.org/10.1007/s00146-010-0270-x

Vallée-Tourangeau, G., Abadie, M., & Vallée-Tourangeau, F. (2015). Interactivity fosters Bayesian reasoning without instruction. *Journal of Experimental Psychology: General*, *144*(3), 581–603. https://doi.org/10.1037/a0039161

Vallée-Tourangeau, G., & Vallée-Tourangeau, F. (2017). Cognition beyond the classical information processing model: Cognitive interactivity and the systemic thinking model (SysTM). In S.J Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain* (pp. 133–155). Springer. https://doi.org/10.1007/978-3-319-49115-8

Valueva, E., Lapteva, E., & Ushakov, D. (2016). Aha-cueing in problem solving. *Learning and Individual Differences*, *52*, 204–208. https://doi.org/10.1016/j.lindif.2016.02.003

van Dijk, J., Kerkhofs, R., van Rooij, I., & Haselager, P. (2008). Can there be such a thing as embodied embedded cognitive neuroscience? *Theory & Psychology*, *18*(3), 297–316. https://doi.org/10.1177/0959354308089787

Van Gelder, T. (1995). What might cogniton be if not perception. *The Journal of Philsophy*, *41*(7), 345–381. https://doi.org/10.2307/2941061

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*, 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Varela, F. J., Thompson, E., & Rosch, E. (2016). *The embodied mind: Cognitive science and human experience* (revised edition). MIT Press.

Wallace, B. (2007). Introduction. In B. Wallace, A. Ross, J. Davies, & T. Anderson (Eds.), *The mind, the body, and the world: Psychology after cognitivism?* (pp. 1–33). Imprint Academic.

Wallas, G. (1926). *The art of thought*. Jonathan Cape.

Webb, M. E., Little, D. R., & Cropper, S. J. (2016). Insight is not in the problem: Investigating insight in problem solving across task types. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01424

Webb, M. E., Little, D. R., & Cropper, S. J. (2019). *Individual differences in insight: An alternative (uses) task.* https://doi.org/10.13140/RG.2.2.14411.13605

Webb, M. E., Little, D. R., & Cropper, Simon. J. (2018). Once more with feeling: Normative data for the aha experience in insight and noninsight problems. *Behavior Research Methods*, *50*(5), 2035–2056. https://doi.org/10.3758/s13428-017-0972-9

Webb, S., & Vallée-Tourangeau, F. (2009). Interactive word production in dyslexic children. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Shomaker (Eds.), *Proceedings of the Thirty-first Annual Conference of the Cognitive Science Society.* (pp. 1436–1441). Cognitive Science Society.

Weisberg, R. (2010). The study of creativity: From genius to cognitive science. *International Journal of Cultural Policy*, *16*(3), 235–253. https://doi.org/10.1080/10286630903111639

Weisberg, R. W. (1986). *Creativity: Genius and other myths*. W.H. Freeman.

Weisberg, R. W. (1995). Prolegomena to theories of insight in problem solving: A taxonomy of problems. In R. J. Sternberg & J. Davidson (Eds.), *The nature of insight* (pp. 157–196). MIT Press.

Weisberg, R. W. (2015). Toward an integrated theory of insight in problem solving. *Thinking & Reasoning*, *21*(1), 5–39. https://doi.org/10.1080/13546783.2014.886625

Wetherell, M. (2007). A step too far: Discursive psychology, linguistic ethnography and questions of identity. *Journal of Sociolinguistics*, *11*(5), 661–681. https://doi.org/10.1111/j.1467-9841.2007.00345.x

Wheeler, M. (2004). Is language the ultimate artefact? *Language Sciences*, *26*(6), 693–715. https://doi.org/10.1016/j.langsci.2004.09.011

Wheeler, M. (2020). *Cognition in the (not so) world*. Virtual Psychology, Philosophy and Informatics Group. https://www.youtube.com/watch?v=tXEr4IkWfbs&feature=youtu.be&app=desktop

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal fluency: Language or executive function measure? *Applied Neuropsychology: Adult*, *23*(1), 29–34. https://doi.org/10.1080/23279095.2015.1004574

Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00058

Wilson, R. A. (2001). *Boundaries of the mind: The individual in the fragile sciences*. Cambridge University Press.

Wilson, R. A. (2014). Ten questions concerning extended cognition. *Philosophical Psychology*, *27*, 19–33. https://doi.org/10.1080/09515089.2013.828568

Wilson, R. A., & Clark, A. (2009). How to situate cognition. In P. Robbins & M. Ayede (Eds.), *The Cambridge handbook of situated cognition* (pp. 55–78). Cambridge University Press.

Yaqub, O. (2018). Serendipity: Towards a taxonomy and a theory. *Research Policy*, *47*, 169–179. https://doi.org/10.1016/j.respol.2017.10.007

Yin, R. K. (2014). *Case study research: Design and methods* (Fifth edition). SAGE.

Zagaria, A., Ando', A., & Zennaro, A. (2020). Psychology: A giant with feet of clay. *Integrative Psychological and Behavioral Science*, *54*(3), 521–562. https://doi.org/10.1007/s12124-020-09524-5