Cognitive and evolutionary foundations of culture and belief

Robert Malcolm Ross MSc

ARC Centre of Excellence in Cognition and its Disorders

Department of Cognitive Science

Macquarie University, Sydney, Australia

A thesis submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

August 2015

Table of Contents

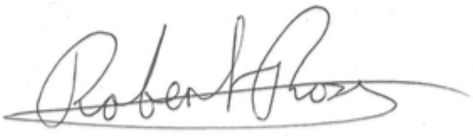
Summary

This thesis explores topical issues in human culture and belief using tools afforded by cognitive psychology and evolutionary theory. Chapter 1 outlines the specific topics examined in this thesis. Chapter 2 presents a meta-analysis that examines the association between delusional ideation and data gathering in the “beads task” paradigm. Chapter 3 presents a behavioural study that examines the extent to which analytic cognitive style and delusional ideation independently predict data gathering in the “beads task” paradigm. Chapter 4 presents a behavioural study of belief formation using the “allergist” associative learning paradigm. Chapter 5 presents an analysis of the evolution of European folktales using methods from population genetics to examine cultural evolution in large, modern societies. Chapter 6 presents a discussion of the importance of taking a geographically explicit approach to the analysis of cross-cultural data. Chapter 7 presents an analysis of the evolution of Arctic folktales using methods from population genetics to examine cultural evolution in small, traditional societies. Chapter 8 presents a general conclusion that summarises the contribution that this thesis makes to our understanding of culture and belief.

Declaration

The work in this thesis is my own original work. It has not been submitted for a higher degree in any other university or institution. All of the work reported in this thesis was undertaken during the time I was enrolled as a PhD student at Macquarie University, under the supervision of Associate Professor Robyn Langdon, Emeritus Professor Max Coltheart, and Doctor Ryan McKay. Ethics approval for the studies reported in this thesis was obtained from Macquarie University's Human Research Ethics Committee, Reference No. 5201300621.

Signed:



Robert Malcolm Ross

30th August 2015

Acknowledgements

Thank you to my supervisors Associate Professor Robyn Langdon, Emeritus Professor Max Coltheart, and Doctor Ryan McKay for your mentorship, encouragement, and advice. I’m a better researcher for having studied with you.

Thank you to my other collaborators Associate Professor Quentin Atkinson, Doctor Will Gervais, Doctor Simon Greenhill, Doctor Oren Griffiths, and Gordon Pennycook for providing much needed expertise. And thank you to Lindsay for proofreading.

Thank you to my family—my parents Heather and Duncan for supporting me during my convoluted journey through academia, and my brother John for graciously tolerating big brother time while I was figuring out whether more study was right for me.

Thank you to the people I’ve lived with (or house sat for) since I started my PhD—Nick, Radhika, Parviz, Karthik, Vignesh, Manveen, Aline, Matt, Rach, and Mal.

Thank you to Margrit, Tony, Anne, and Kristy for making me feel welcome in your home, and including me in holidays and family events.

Thank you to the many folks I’ve shared good times (and tough times) with at MACCS and CCD, in NZ, and scatted around the world.

And a special thank you to Dani for providing me with love and support, for keeping me going during the difficult times, and for all the wonderful times we’ve shared.

Chapter One: Introduction

2

1.1 Introduction

Biologically, humans are a very homogeneous species. We exhibit considerably less genetic variation than our closest cousin the chimpanzee, for example. By contrast, in terms of culture and belief we are a spectacularly diverse species, exhibiting variation at an extraordinary scale seen nowhere else in the natural world. We speak thousands of mutually incomprehensible languages—but can become fluent speakers of any of these languages if exposed to them from a young enough age. Some of us hunt, others forage, while still others cultivate the land—diverse approaches to provisioning that require remarkably complex skills that have taken generations to perfect. We worship a vast array of different gods and perform elaborate rituals to win their favour—rituals that seem bizarre to people from other cultures, but perfectly natural to us. Such profound diversity attests to the fact that our complex cognitive system is surely uniquely evolved for the transmission of culture and belief. Unfortunately, this cognitive system can fail, resulting in maladaptive beliefs and behaviours.

In this thesis I examine topical questions pertaining to belief and culture. The thesis is divided into two sections that approach these topics from different perspectives. In the first section, I use method and theory from cognitive psychology to examine cognitive mechanisms involved in belief. In particular, I examine information processing biases and deficits that scholars have proposed play an important role in the formation and maintenance of delusions. In the second section, I use method and theory adapted from population genetics to examine cultural change from an evolutionary perspective. In particular, I examine the transmission of folktales within and between ethnolinguistic groups.

Prior to exploring these issues, in this introductory chapter I outline theories and controversies that motivate the research presented in this thesis. I begin with a brief summary of contemporary discussions about the nature of belief and rationality. This is followed by a defence of the hypothesis that delusions are a species of irrational belief. Next, I defend two contentious claims about delusional beliefs: that they lie on a continuum with other beliefs, and that cultural context plays a role in determining which particular beliefs are delusions. This discussion of cultural context leads me onto a sketch of evolutionary approaches to the study of cultural change and my suggestion that these approaches hold considerable potential for examining the relationship between belief, culture, rationality, and delusions. Finally, I outline the specific research context and rationale for the studies discussed in the chapters that follow.

1.2 Belief

The term "belief" is generally used to refer to "the attitude we have, roughly, whenever we take something to be the case or regard it as true" (Schwitzgebel, 2015). Over the past several decades, sophisticated models of many core cognitive processes have been developed (e.g., attention, memory, perception, language, problem solving); nevertheless, cognitive models of belief remain relatively underspecified (Connors & Halligan, 2015). This lag is understandable. Belief is a particularly difficult cognitive process to study scientifically because it draws information from multiple sources and involves many sub-processes. Nevertheless, progress is being made and recently there has been a steady growth of scientific research on belief (e.g. Connors & Halligan, 2015; Galbraith, 2015; Krueger & Grafman, 2013). Much of the progress that is being made is built on insights from research on the cognitive psychology of reasoning that I will explore in this thesis. Fundamental to this reasoning literature is the notion of rationality.

1.3 Rationality

Rationality is typically treated as a normative ideal and can be indexed by the extent to which a belief or behaviour conforms to the optimum defined by normative models[1]. Decades of

[1] I should briefly note that the normative approaches to rationality have been challenged (e.g. Evans & Elqayam, 2011). Nevertheless, anti-normativism is minority position in cognitive science. An attempt to rebut anti-normativism would distract from the focus of this thesis. Consequently, I will follow Stanovich (2011a) and simply stipulate that, "I do not think that

research has demonstrated that typical responses on a variety of reasoning tasks often deviate substantially from normative responses (Kahneman, 2011; Stanovich, 2011b). For instance, in solving syllogistic problems in which believability is in conflict with deductive logic, people frequently provide non-normative responses that are believable, even when specifically instructed to ignore believability and focus on the logical form of the task. In addition, there is a wealth of evidence for individual differences: across a variety of reasoning tasks some people tend to provide normative responses, while others do not (Stanovich, 2011b). Powerful explanations for both the frequency of deviations from normative ideals and individual variation in normative responding are provided by "dual-process" theories of reasoning. According to dual-process theories, the human mind is equipped with two reasoning processes that are relatively distinct (Evans, 2010; Evans & Stanovich, 2013; Kahneman, 2003, 2011; Stanovich, 2011b). The first, Type 1, or "intuitive," processing does not require working memory and provides fast intuitive responses that guide most behaviour. The second, Type 2, or "analytic," processing requires working memory, is relatively slow, and monitors the outputs of Type 1 processing to offer alternative responses.

If we treat performance in laboratory-based reasoning tasks as a microcosm of belief formation in everyday life, then studying departures from normative responding can play an important role in the scientific study of belief. One particularly promising avenue for research is the examination of cognitive biases and dysfunctions that result in abnormal beliefs (Connors & Halligan, 2015; Galbraith, 2015). To this end, recent developments in the literature suggest that the cognitive style of people with delusions might be characterised by an over-reliance on Type 1 processing (Aimola Davies & Davies, 2009; Freeman, Evans, & Lister, 2012; Freeman & Garety, 2014; Freeman, Lister, & Evans, 2014; Garety et al., 2015; Speechley, Murray, McKay, Munz, & Ngan, 2010; Speechley & Ngan, 2008; Speechley, Whitman, & Woodward, 2010; Speechley, Woodward, & Ngan, 2013). That is to say, an

---

psychology will heed any of these admonitions to refrain from normative language… In our lab, subjects *do* make, what we unabashedly call – errors." (p. 270) "intuitive cognitive style" (Stanovich, 2011b) might play a role in the formation and maintenance of delusions. Prior to exploring this possibility (in Chapter 3), it is useful to clarify the nature of delusions.

1.4 Delusions

Much contemporary research in psychiatry and psychology is built on the assumption that delusions are some species of irrational belief. Nevertheless, specifying precisely what additional properties irrational beliefs must have to make them delusions presents a serious challenge. Consider the definition of delusion from the fifth edition of the Diagnostics and Statistical Manual of Mental Disorders (DSM-5):

> Delusion. A false belief based on incorrect inference about external reality that is firmly held despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary. The belief is not ordinarily accepted by other members of the person's culture or subculture (i.e., it is not an article of religious faith). (p. 819)[2]

The DSM definition states that a) delusions are beliefs and that b) a belief can be classified as a delusion only if there is overwhelming evidence that the belief is not true—a clear, if possibly overly stringent, irrationality criterion. Despite the considerable influence of the DSM (it is frequently characterised as "the Bible of psychiatry"), the DSM definition of delusion is very controversial. Coltheart (2007) has succinctly summarised five controversial features of this definition:

> 1. Couldn't a true belief be a delusion, as long as the believer had no good reason for holding the belief? 2. Do delusions really have to be *beliefs*—might they not instead be imaginings that are mistaken for beliefs by the imaginer? 3. Must all delusions be

[2] This is the definition of delusion found in the glossary of terms. Bortolotti (2013) points out that the definition in the section on schizophrenia differs in some important respects. For instance, in the schizophrenia section there is not claim that the belief must be false. Nevertheless, the definition presented in the glossary is more detailed and widely discussed, so I will treat it as "the" DSM definition here.

based on inference? 4. Aren't there delusions that are not about external reality? 'I have no bodily organs' or 'my thoughts are not mine but are inserted into my mind by others' are beliefs expressed by some people with schizophrenia, yet are not about external reality; aren't these nevertheless still delusional beliefs? 5. Couldn't a belief held by all members of one's community still be delusional? (p. 1043)

I agree that some features of the DSM definition are highly problematic. In particular, I am not aware of any careful defences of the claims that delusions must be false (feature 1), delusions must be based on inference (feature 3), or delusions must be about external reality (feature 4). Consequently, it would appear that these features of the DSM definition are not scientifically useful (and are of questionable usefulness for clinicians too I suspect). Nevertheless, I believe that the two other features from Coltheart's list warrant careful consideration because they each play an important role in contemporary debates about the nature of delusions. I will refer to these features of the DSM definition as "delusions as beliefs" (feature 2) and "delusions as culturally relative" (feature 5), and discuss each in turn below.

1.4 Delusions as beliefs

It has been argued that delusions are not, in fact, beliefs (Berrios, 1991; Currie, 2000; Gallagher, 2009; Matthews, 2013; Stephens & Graham, 2004). Among scholars defending this position, there is considerable diversity with respect to positive proposals about the true nature of delusions (Bortolotti 2013). For example, it has been proposed that delusions are imaginings that are misidentified by their subjects as beliefs (Currie, 2000), delusions relate to a parallel reality (Gallagher, 2009), and delusions are empty speech acts that do not carry meaning (Berrios, 1991). Nevertheless, there is considerable overlap with respect to negative arguments that purport to show that delusions are not beliefs. Bortolotti (2013) has provided a useful typology of three clusters of anti-doxastic[3] arguments:

> 1. Beliefs are integrated with other beliefs. If delusions are not integrated with a person's beliefs, then they are not beliefs. 2. Beliefs are responsive to evidence. If delusions are not responsive to evidence, then they are not beliefs. 3. Beliefs guide action. If delusions do not guide action, then they are not beliefs.

Bortolotti (2009) provides detailed rebuttals to these anti-doxastic arguments. Her overarching strategy is to demonstrate that anti-doxastic arguments appeal to exaggerated and idealised claims about the sheer irrationality of delusions relative to other beliefs. Drawing on empirical evidence from psychiatry and psychology, she makes a highly persuasive case that delusions are considerably less irrational than anti-doxastic arguments suggest, and non-delusional beliefs are considerably more irrational than many scholars appreciate. Furthermore, Bortolotti's arguments are supported by the fact that empirical research that treats delusions as beliefs appears to be responsible for important developments in characterising and treating cognitive biases and deficits thought to be involved in clinically significant delusions (Connors & Halligan, 2015; Galbraith, 2015; Garety & Freeman, 2013). Particularly convincing evidence for a doxastic interpretation of delusions is provided by neuropsychological case studies showing that deficits to cognitive systems involved in belief formation seem to play a central role in a variety of monothematic delusions[4] (Coltheart, Langdon, & McKay, 2011). In summary, contemporary scholarship provides evidence that existing anti-doxastic arguments should not dissuade scholars from conducting empirical research that provisionally assumes that many, perhaps all, delusions are beliefs.

1.5 Delusions as continuous with non-delusional beliefs

---

[3] "Doxastic" means "pertaining to belief". Doxastic theories of delusions claim that delusions are beliefs, and anti-doxastic theories claim that delusions are not beliefs.

[4] An individual with a "monothematic delusion" holds a single delusional belief, or a small set of delusional beliefs pertaining to a single theme. An individual with "polythematic delusions" hold multiple delusional beliefs that do not have a unifying theme. Although this distinction is not without difficulties (Radden, 2013), it does appear to be very useful (Coltheart, 2013).

An important issue not addressed by the DSM definition of delusion is whether clinical delusions are continuous with unusual beliefs in the general population, or if they are a qualitatively distinct category of belief (DeRosse & Karlsgodt, 2015). A large body of empirical research suggests that attenuated forms of positive symptoms of psychosis (hallucinations and delusions) are not uncommon in the general population. A recent meta-analysis of 61 cohorts found a median annual incidence of delusions of 1.5% and a prevalence rate of 4.9% (Linscott & van Os, 2013). Furthermore, meta-analyses provide evidence that the presence of positive symptoms in non-clinical populations predicts later psychotic illness (Fusar-Poli et al., 2012; Kaymaz et al., 2012) and that early intervention might, in some cases, delay or prevent transition to psychosis (Stafford, Jackson, Mayo-Wilson, Morrison, & Kendall, 2013). Such research is often interpreted as providing strong evidence for some manner of continuity between normality and abnormality. For example, Bentall (2003) proposes a "principle of continuity":

> Abnormal behaviours and experiences are related to normal behaviours and experiences by continua of frequency (the same behaviours and experiences occur less frequently in non-psychiatric populations), severity (less severe forms of the behaviours and experiences can be identified in non-psychiatric populations), and phenomenology (non-clinical analogues of the behaviours and experiences can be identified as part of normal life). (p. 115)[5]

Nevertheless, a number of scholars have argued that we should not endorse continuity arguments too hastily (David, 2010; Lawrie, Hall, McIntosh, Owens, & Johnstone, 2010; Sommer, 2010). One of the most important objections is that there are problems with the empirical evidence used to support claims that "delusion-like" (i.e. not clinically significant) beliefs in the general population are not quantitatively different to "true" (i.e. clinically

[5] This principle of continuity could be worded more carefully to include beliefs since beliefs are not behaviours or experiences per se.

significant) delusional beliefs. For instance, consider the "Peters *et al*. Delusions Inventory" (PDI; Peters, Joseph, Day, & Garety, 2004; Peters, Joseph, & Garety, 1999), which is the most widely used measure of delusion-like beliefs designed for use in nonclinical populations. In the PDI, terms such as "feel" and "as if" are added to questions that are used in clinical diagnostic contexts. For instance, one item from the PDI asks, "Do you ever feel as if people are reading your mind?" Consequently, it is not clear that the PDI (and other measures that are worded in similar ways) probes beliefs as opposed to experiences or imaginings. I agree that this issue—and other issues raised by critics of the continuity thesis—are of genuine concern. Nevertheless, there is evidence that scores on measures of delusion-like belief, such as the PDI, are associated with reasoning biases that are thought to play a role in delusion formation. For instance, there is evidence that a "jumping to conclusions" cognitive bias is associated with both delusion severity in clinical populations and PDI scores in the general population (Garety & Freeman, 1999, 2013). Such results provide prima facie evidence that it is worthwhile to continue to investigate the continuity hypothesis despite the apparent limitations of existing measures of non-clinical delusions.

1.6 Delusions as culturally relative

Scholars have also questioned whether cultural context should play a role in determining whether a belief should be classified as a delusion:

> Does the belief have to be different from what almost everyone else believes? If a bizarrely implausible belief is formed and sustained in ways that are characteristic of delusions, then it seems that, for the purpose of psychological theory, it should be grouped together with delusions even if many other subjects believe the same thing. (Davies, Coltheart, Langdon, & Breen, 2001, p. 134)

The striking example of Koro—or "penis panic" as it is sometimes known—has been used to highlight the tension (my italics):

> Koro is a condition in which a person believes that his or her sexual organs (penis in men, vulva or nipples in women) are being retracted into the body, and that death will follow from full retraction. There are many reports from South-East Asia of this belief spreading rapidly throughout whole subcultures. For example, in just 10 days in 1967 in Singapore, 469 cases of this condition arose (Ngui, 1969); and in June to September 1982 there was an epidemic of this belief in northeastern India, in three districts of Assam and two nearby districts of Bengal, to the extent that "for a few weeks, the whole area was in the grip of a fear of the illness" (Sachdev, 1985, p. 434). *It is hard to see what justification there could possibly be for asserting that Koro should not count as a delusional condition, yet that assertion follows from the DSM-IV definition.* (Coltheart, in press)

The suggestion here seems to be that, at least in some cases, deviations from rational evaluations of evidence can be so extreme that cultural context cannot be plausibly offered as an exemption, and the belief should be classified as a delusion.

By contrast, other scholars have argued that there are cases in which it is perfectly legitimate to excuse a belief as not being delusional because the belief is culturally sanctioned. This position has been forcefully defended by Murphy:

> Normal human cognitive development includes the acquisition of beliefs and other mental states, not just from the world, but from other people. If we come across individuals who have beliefs that are important to them but seem to be based on no causal contact with the outside world, nor on testimony, we are entitled to wonder about them, especially if no rational justification for the belief can be given. So even if you believe, as I do, that rational justification for religious belief will not be forthcoming, it is undeniable that normal maturing brains do pick up religion, along with many other false theories about the world. That, in fact, is why religion is not delusional even if religious beliefs are false. Our evolved psychology, it appears, just

> commits us to the same epistemic mistakes, generation after generation. (Murphy, 2006, p. 181)
>
> It's normal for people to pick up beliefs that we find weird from the culture around them, and not normal for them to arrive at equivalently weird beliefs all by themselves in cultures that provide no support for such beliefs. (Murphy, 2013, p. 119)

Murphy supports his position with a striking example of his own (my italics):

> Boyer (2001, p. 69–70), reporting fieldwork done by Wendy James in the Sudan, discusses ebony trees that are believed to be a source of social information. The trees record conversations, and are privy to the plans of witches. You can learn what they know by burning an ebony twig, dipping it in water and reading the pattern of ashes in the water. *A belief in cognitive interaction with ebony trees counts as culturally normal, and hence not delusional or otherwise suspect.* This is an article of the local religion, or more broadly of local beliefs about the working of the universe and their significance for human life. (Murphy, 2013, p. 118-119)

Murphy contrasts this Sudanese example with a (true) case of a man from a Western cultural background who had experiences of trees talking to him (i.e. a man with auditory hallucinations) and, as a result, came to believe that trees can talk. Here Murphy suggests that we would not hesitate to classify this individual as holding a delusional belief since this belief is in no way supported by his culture. As Murphy succinctly puts it, "numbers matter with delusions" (2006, p. 182)[6]. Similarly, in discussing how we should go about developing a cognitive model of delusions, Langdon (2013) has argued that,

> The model of normal belief formation that we apply to explain delusions ought to

[6] Murphy uses his argument for cultural relativism to defend a theory that delusions should be defined, predominantly, in terms of folk psychology and folk epistemology, not in terms of a cognitive model of belief formation (Murphy, 2006, 2011, 2013). According to Murphy, "delusions are beliefs that we cannot explain in any folk psychological terms" (2011, p. 19). Although I am highly sympathetic to Murphy's defence of the cultural relativism of delusions, I do not endorse his folk theory of delusions. See Bortolotti (2011) and Radden (2013) for brief, but persuasive, summaries of problems associated with Murphy's folk theory.

> include the normal processes for socially transmitted beliefs. These normal processes include a default to believe what respected other people, and previously reliable sources for beliefs, tell is true. (p. 80)

I agree with Langdon. Furthermore, as I discuss in Section 1.7, I suggest that modern cultural evolutionary theories hold considerable potential for examining the cultural contexts under which beliefs that seem bizarre in a Western context—such as belief in Koro and belief in talking trees—might not, in fact, be delusional.

As Murphy has suggested, less exotic examples also provide support for cultural relativism. For the non-religious, it can be tempting to conclude that many (if not all) religious beliefs are, in fact, delusions. That is precisely what Richard Dawkins argues in *The God Delusion* (Dawkins, 2006). However, this move creates at least two serious problems. First, there are practical issues pertaining to clinical diagnosis and treatment. Although a scientific definition of delusions should not be entirely beholden to clinical utility, I think that there should be some relationship been scientific and clinical concerns. If most of humanity can be categorised as delusional because they hold religious beliefs, the presence of delusional beliefs provides little guidance in determining who might benefit from treatment from psychiatrists and psychologists. Second, scholarship on delusions typically take delusions to be, in some sense, manifestations of underlying *dysfunction*—that is to say, deviation from normal functioning. However, as Murphy notes, the sheer prevalence of religious belief (religious belief appears to be a cross-cultural universal; Brown, 1991) seems to provide compelling evidence that the *normally functioning* human cognitive system frequently picks up religious belief. Furthermore, many scholars argue that religion is, in fact, an evolutionary adaptation—that is to say, the human cognitive system has been *designed by natural selection* to pick up religious belief (Bering, 2011; Purzycki & Sosis, 2009; Wilson, 2002; but see McKay & Dennett, 2009).

In summary, I suggest that adopting a definition of delusion that does not take cultural context into account risks creating an almost unrecognisable definition that lacks practical applications. It is not uncommon for writers to use culture-free definitions of delusion for rhetorical effect, but it is far from clear that such definitions are useful for psychological research (my italics):

> The best way of checking the reality of our perceptions is to confirm that they correspond to the experiences of everyone else. This mutual checking applies not only to what we experience with our senses, but also to those things we believe about the world that are not based directly on our senses. *Our view of reality can be conceived of as a mass delusion*; 500 years ago we 'knew' that the Earth was flat, now we know it is round. (Frith & Johnstone, 2003, p. 160)

Murphy's defence of the cultural relativity of delusions is not uncontroversial. For example, Radden (2013) has expressed some scepticism about its utility,

> Within today's cultures and subcultures, however, it seems to me there is too much 'noise' for Murphy's [religious and cultural exemption] to be a useful guide in sorting culturally acceptable weird ideas. (p. 129)

This scepticism is understandable. In modern pluralistic societies we are exposed to a tremendous variety of belief systems, and there is considerable scope for us choose which subculture we would like to belong to. Nevertheless, I suggest that the issue of whether or not contemporary culture is "too noisy" is an *empirical* question that warrants careful investigation. To date, there is little empirical research that confronts this issue directly. In this thesis, I do not attempt to confront this issue directly either. I think that a better strategy is for empirical research to focus on mapping out the parameters of normal cultural transmission before examining just how far removed from normal cultural transmission a belief must be before it could qualify as being delusional. Consequently, one of the goals of this thesis is to examine normal cultural transmission.

1.7 Cultural change as an evolutionary process

Scholars have long noted remarkable parallels between biological and cultural evolution (Mesoudi, 2011), including Charles Darwin himself (Mesoudi, 2004). For example, in the case of language, Darwin has pointed out that,

> The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process are curiously parallel… We find in distinct languages striking homologies due to community of descent, and analogies due to similar processes of formation. (Darwin, 1871, p. 90-91)

And William James, one of the founders of modern psychology, noted that,

> A remarkable parallel… obtains between the facts of social evolution on the one hand, and the zoological evolution as expounded by Mr. Darwin on the other. (James, 1880, p. 441)

Despite the parallels between biological and cultural change having long been recognised by such esteemed scholars as Darwin and James, it has been only recently that serious progress has been made in developing a rigorous quantitative science of cultural evolution (Mesoudi, 2011; Richerson & Boyd, 2005). According to "dual-inheritance theory" (also known as gene-culture co-evolutionary theory), genes are not the only inheritance system; culture acts as an inheritance system too (Boyd & Richerson, 1985; Cavalli-Sforza & Feldman, 1981; Mesoudi, 2011; Richerson & Boyd, 2005). That is to say, like biological change, cultural change "encompasses Darwinian processes that include variation, competition, selection, inheritance, accumulation of modifications, adaptation, geographical distribution, convergent evolution and changes of function" (Whiten, Hinde, Laland, & Stringer, 2011, p. 939-940).

A major focus of research in the dual-inheritance theory literature has been to explain population-level cultural evolutionary processes by adapting evolutionary models and statistical techniques developed in population genetics (Boyd & Richerson, 1985; Cavalli-Sforza & Feldman, 1981; Mesoudi, 2011; Richerson & Boyd, 2005). "Population" is a notoriously difficult concept to define; nevertheless, in human research, populations are typically operationalised as self-identified "ethnolinguistic groups" (groups defined by shared language and cultural identity). It is not without good reason that ethnolinguistic groups are a major focus of research. There is strong evidence for the existence of cultural evolutionary processes that inhibit the transmission of cultural information between groups and facilitate the transmission of cultural information within groups (Mesoudi, 2011; Richerson & Boyd, 2005). In fact, an influential theory argues that humans are particularly sensitive to markers of group identity because this facilitates cooperation within groups for competing against other groups (Richerson et al., in press; but see West, El Mouden, & Gardner, 2011).

Dual-inheritance theory recognises a variety of cultural evolutionary forces. Richerson and Boyd (2005, p. 69) outline what is probably the most influential typology. Among the evolutionary forces in their typology, the "transmission biases" are particularly plausible candidates for cultural evolutionary processes that could validate the religious and cultural exemptions defended by Murphy (see Section 1.6). Firstly, "content-based biases"—biases to retain cultural variants whose content has a "good fit" with the human mind. Being easily remembered, for example. Secondly, "frequency-biased biases"—biases for more common cultural variants to be retained. Popular cultural variants, for example. Thirdly, "model-based biases"—biases for the traits of individuals who carry cultural variants to influence whether a cultural variant is retained. Paying attention to prestigious individuals, for example. A diverse body of empirical research provides strong evidence that these transmission biases have an important influence on the relative frequency of different cultural traits in human populations (Mesoudi, 2011; Richerson & Boyd, 2005), including group-level cultural and religious beliefs (Richerson et al., in press; Norenzayan et al., in press). In Section 1.6, I drew attention to a claim that cognitive models of delusions need to account for normal processes of socially transmitted beliefs (Langdon, 2013), but also to scepticism about the possibility of using religious and cultural context to identify which "weird" beliefs are not delusions (Radden, 2013). I suggest research from the perspective of dual-inheritance theory has the potential to make important contributions to this discussion.

1.8 Thesis overview

1.8.1 Section one: delusional belief and rationality

Research suggests that delusional and delusion-prone individuals gather less information before forming beliefs than controls. This "jumping to conclusions" (JTC) bias has been presented as one of the most strongly supported cognitive biases associated with delusions (Garety & Freeman, 2013). Much of the evidence for the existence of the JTC bias comes from studies using the "beads task" data-gathering paradigm that was introduced into the delusions literature in a pioneering study by Huq, Garety, and Hemsley (1988). Chapter 2 and Chapter 3 report complementary investigations of the beads task.

Chapter 2 is based on a recent paper published in *Schizophrenia Bulletin* (Ross, McKay, Coltheart, & Langdon, 2015)[7] that reports a meta-analysis investigating the association between data gathering in the beads task and delusional ideation indexed using the PDI. We decided to undertake a meta-analysis because previous assessments of the beads task literature—two systematic reviews (Garety & Freeman, 1999, 2013) and a meta-analysis (Fine, Gardner, Craigie, & Gold, 2007)—had significant limitations (discussed in Chapter 2).

Chapter 3 is based on a recently submitted manuscript (Ross, Pennycook, et al., submitted) that presents an empirical study that examines predictors of data gathering in the beads task. Specifically, we investigate the novel hypothesis that increased "analytic cognitive style" (see Section 1.3) predicts greater data gathering in the beads task. In addition, we aim to replicate previous studies that have suggested that greater PDI scores predict less data gathering (Garety & Freeman, 2013).

Chapter 4 reports a study that uses the "allergist task" associative learning paradigm

[7] This paper won the 2014 ARC Centre of Excellence in Cognition and its Disorders Excellence in Research Student Award.

(Dickinson & Burke, 1996) to extend earlier research that has used this task to study delusions. A recent brain imaging study by Corlett, Murray, *et al.* (2007) reported evidence that a population diagnosed with psychosis display abnormal patterns of brain activation in right lateral prefrontal cortex that are associated with learning deficits in the allergist task. Importantly, the measure of "unusual thought content" from the Brief Psychosis Ratings Scale (Ventura, Shaner, & Lieberman, 1993) was found to be the strongest predictor of abnormal neural activation. Results from this study have been interpreted as providing tentative support for cognitive theories of delusional belief that implicate dysfunction in right dorsolateral prefrontal cortex (Coltheart et al., 2011). Nevertheless, careful examination of the protocol employed by Corlett *et al.* (2007) reveals limitations that render the behavioural evidence presented in this study difficult to interpret (Griffiths, Langdon, Le Pelley, & Coltheart, 2014; Griffiths, Le Pelley, & Langdon, 2015; but see Corlett & Fletcher, 2015). We develop a closely matched behavioural paradigm that tests the behavioural evidence for learning more rigorously.

1.8.2 Section two: transmission of culture

The three chapters in section two focus on the analysis of individual-level and group-level ethnolinguistic data using method and theory from population genetics and dual-inheritance theory.

Chapter 5 is based on a recent paper published in *Proceedings of the Royal Society* (Ross, Greenhill, & Atkinson, 2013)[8]. This study examines 700 variants of a folktale in 31 European ethnolinguistic groups to quantify the extent to which geographic distance and

---

[8] This paper won the 2013 ARC Centre of Excellence in Cognition and its Disorders Excellence in Research Student Award, and generated some media interest:
Business Insider – "Amazing folktale map reveals deep links between cultures", 20 December 2014
National Geographic – "Humans swap DNA more readily than they swap stories", 6 February 2013
Nature – "Genes mix faster than stories", 06 February 2013
New Scientist – "Genes mix across boarders more easily than folk tales", 06 February 2013

cultural barriers account for the distribution of variation in narrative elements of this folktale. The primary aim is to map out the parameters of normal cultural transmission for a trait that is “selectively neutral” (folktales, unlike many other cultural inventories, such as tools, are not tested against the physical environment and therefore do not have a direct effect on survival).

Chapter 6 is based on an invited commentary soon to be published in *Behavioral and Brain Sciences* (Ross & Atkinson, in press). This commentary critically evaluates empirical evidence presented in a target article to support the theory of cultural group selection (Richerson et al., in press). According to cultural group selection, Darwinian selection does not occur only at the level of the gene, but also at the level of the group too. This theory is not uncontroversial (West et al., 2011), and one important criticism is that it is not clear that there is sufficient stable group-level structure for Darwinian forces to operate. One of the most exciting features of the target article is that it reports new quantitative evidence for cultural differentiation between groups (see section 4.1 of Richerson et al., in press). However, in Chapter 6, we argue that there is a problem with the evidence they present: cultural differentiation does not entail group-level structure. We use our cross-cultural analysis of folktales (Ross et al., 2013) and cross-cultural analyses of other cultural inventories as empirical evidence.

Chapter 7 is based on a manuscript soon to be published in Evolution and Human Behavior (Ross & Atkinson, in press) that analyses folktale diversity in Arctic cultures of Siberia, Alaska, Canada, and Greenland. This chapter follows up on some outstanding questions posed by our analysis of European folktales (Ross et al., 2013). In particular, in our analysis of European folktales we speculate that the patterns we identify can be used to make inferences about general patterns of cultural transmission in humans. However, this argument makes some assumptions. Modern European cultures differ from cultures that characterised much of human evolution in some important respects. Particularly noteworthy is that

European cultures had writing systems, and literacy may have played an important role in the dissemination of folktales in Europe (Bottigheimer, 2009; de Blécourt, 2012). In Chapter 7, we analyse the folktales of low-density hunter-gather groups that do not have a long history of writing. These groups arguably provide a better window into the parameters of normal cultural transmission during human prehistory.

1.9 Discussion and conclusion

Finally, in Chapter 8, I integrate insights from these chapters and discuss the contribution this thesis makes to scholarship on the cognitive and evolutionary foundations of human culture and belief.

1.10 References

Aimola Davies, A. M., & Davies, M. (2009). Explaining pathologies of belief. In M. R. Broome & L. Bortolotti (Eds.), *Psychiatry as cognitive neuroscience: philosophical perspectives* (pp. 285-323).

Bentall, R. P. (2003). *Madness explained: psychosis and human nature*. London, UK: Penguin.

Bering, J. (2011). *The belief instinct: the psychology of souls, destiny, and the meaning of life*. New York, NY: W. W. Norton & Company.

Berrios, G. E. (1991). Delusions as 'wrong beliefs': a conceptual history. *British Journal of Psychiatry, 159*(suppl. 14), 6-13.

Bortolotti, L. (2009). *Delusions and other irrational beliefs*. Oxford, UK: Oxford University Press.

Bortolotti, L. (2011). In defence of modest doxasticism about delusions. *Neuroethics*. doi: 10.1007/s12152-011-9122-8

Bortolotti, L. (2013). Delusion. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/delusion/.

Bottigheimer, R. B. (2009). *Fairy tales: a new history*. Albany, NY: State University of New York.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago, Il: Chicago University Press.

Boyer, P. (2001). *Religion explained: the evolutionary origins of religious thought*. Basic Books: New York, NY.

Brown, D. E. (1991). *Human universals*. New York, NY: McGraw-Hill.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: a quantitative approach*. Princeton, NJ: Princeton University Press.

Coltheart, M. (2007). The 33rd Sir Frederick Bartlett Lecture: cognitive neuropsychiatry and delusional belief. *The Quarterly Journal of Experimental Psychology, 60*(8), 1041-1062. doi: 10.1080/17470210701338071

Coltheart, M. (2013). On the distinction between monothematic and polythematic delusions. *Mind & Language, 28*(1), 103-112. doi: 10.1111/mila.12011

Coltheart, M. (in press). Delusions. In R. Scott & S. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences*. Hoboken, NJ: John Wiley & Sons.

Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional belief. *Annual Review of Psychology, 62*(5), 271-298. doi: 10.1146/annurev.psych.121208.131622

Connors, M. H., & Halligan, P. W. (2015). A cognitive account of belief: a tentative roadmap. *Frontiers in Psychology, 5*(1588), 1-14. doi: 10.3389/fpsyg.2014.01588

Corlett, P. R., & Fletcher, P. C. (2015). Delusions and prediction error: clarifying the roles of behavioural and brain responses. *Cognitive Neuropsychiatry*, *20*(2) 95-105. doi: 10.1080/13546805.2014.990625

Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R. F., Shanks, D. R., Robbins, T. W., . . . Fletcher, P. C. (2007). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain, 130*, 2387-2400. doi: 10.1093/brain/awm173

Currie, G. (2000). Imagination, delusion and hallucination. *Mind & Language, 15*(1), 168-183. doi: 10.1111/1468-0017.00128

Darwin, C. (1871). *The descent of man*. London, UK: Gibson Square, 2003. Original edition, 1871.

David, A. S. (2010). Why we need more debate on whether psychotic symptoms lie on a continuum with normality. *Psychological Medicine, 40*(12), 1935-1942. doi: 10.1017/S0033291710000188

Davies, M., Coltheart, M., Langdon, R., & Breen, N. (2001). Monothematic delusions: towards a two-factor account. *Philosophy, Psychiatry, & Psychology, 8*(2), 133-158. doi: 10.1353/ppp.2001.0007

Dawkins, R. D. (2006). *The god delusion*. London, UK: Transworld.

de Blécourt, W. (2012). *Tales of magic, tales in print: on the genealogy of fairy tales and the Brothers Grimm*. Manchester, UK: Manchester University Press.

DeRosse, P., & Karlsgodt, K. H. (2015). Examining the psychosis continuum. *Current Behavioral Neuroscience Reports, Published online ahead of print*. doi: 10.1007/s40473-015-0040-7

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgments. *Quarterly Journal of Experimental Psychology Section B, 49*(1), 60-80. doi: 10.1080/713932614

Evans, J. S. B. T. (2010). *Thinking twice: two minds in one brain*. Oxford, UK: Oxford University Press.

Evans, J. S. B. T., & Elqayam, S. (2011). Subtracting "ought" from "is": descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences, 34*(05), 233-290. doi: 10.1017/S0140525X11001440

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: advancing the debate. *Perspectives on Psychological Science, 8*(3), 223-241. doi: 10.1177/1745691612460685

Fine, C., Gardner, M., Craigie, J., & Gold, I. (2007). Hopping, skipping or jumping to conclusions? Clarifying the role of the JTC bias in delusions. *Cognitive Neuropsychiatry, 12*(1), 46-77. doi: 10.1080/13546800600750597

Freeman, D., Evans, N., & Lister, R. (2012). Gut feelings, deliberative thought, and paranoid ideation: A study of experiential and rational reasoning. *Psychiatry Research, 197*(1-2), 119-122. doi: 10.1016/j.psychres.2011.12.031

Freeman, D., & Garety, P. A. (2014). Advances in understanding and treating persecutory delusions: a review. *Social Psychiatry and Psychiatric Epidemiology, 49*(8), 1179-1189. doi: 10.1007/s00127-014-0928-7

Freeman, D., Lister, R., & Evans, N. (2014). The use of intuitive and analytic reasoning styles by patients with persecutory delusions. *Journal of Behavioral Therapy and Experimental Psychiatry, 45*(4), 454-458. doi: 10.1016/j.jbtep.2014.06.005

Frith, C., & Johnstone, E. C. (2003). *Schizophrenia: a very short introduction*. Oxford, UK: Oxford University Press.

Fusar-Poli, P., Bonoldi, I., Yung, A. R., Borgwardt, S., Kempton, M. J., Valmaggia, L., . . . McGuire, P. (2012). Predicting psychosis: meta-analysis of transition outcomes in individuals at high clinical risk. *Archives of General Psychiatry, 69*(3), 220-229. doi: 10.1001/archgenpsychiatry.2011.1472

Galbraith, N. (Ed.). (2015). *Aberrant beliefs and reasoning*. New York, NY: Psychology Press.

Gallagher, S. (2009). Delusional realities. In M. R. Broome & L. Bortolotti (Eds.), *Psychiatry as cognitive neuroscience* (pp. 245-266). Oxford, UK: Oxford University Press.

Garety, P. A., & Freeman, D. (1999). Cognitive approaches to delusions: a critical review of theories and evidence. *British Journal of Clinical Psychology, 38*, 113-154. doi: 10.1348/014466599162700

Garety, P. A., & Freeman, D. (2013). The past and future of delusions research: from the inexplicable to the treatable. *British Journal of Psychiatry, 203*(5), 327-333. doi: 10.1192/bjp.bp.113.126953

Garety, P. A., Waller, H., Emsley, R., Jolley, S., Kuipers, E., Bebbington, P., . . . Freeman, D. (2015). Cognitive mechanisms of change in delusions: an experimental investigation targeting reasoning to effect change in paranoia. *Schizophrenia Bulletin, 41*(2), 400-410. doi: 10.1093/schbul/sbu103

Griffiths, O., Langdon, R., Le Pelley, M. E., & Coltheart, M. (2014). Delusions and prediction error: re-examining the behavioural evidence for disrupted error signalling in delusion formation. *Cognitive Neuropsychiatry, 19*(5), 439-467. doi: 10.1080/13546805.2014.897601

Griffiths, O., Le Pelley, M. E., & Langdon, R. (2015). The bridge between neuroscience and cognition must be tethered at both ends. *Cognitive Neuropsychiatry*, *20*(2), 106-108. doi: 10.1080/13546805.2014.993464

Huq, S. F., Garety, P. A., & Hemsley, D. R. (1988). Probabilistic judgments in deluded and non-deluded subjects. *Quarterly Journal of Experimental Psychology Section A, 40*(4), 801-812. doi: 10.1080/14640748808402300

James, W. (1880). Great men, great thoughts, and the environment. *Atlantic Monthly, 46*, 441-459.

Kahneman, D. (2003). Maps of bounded rationality: psychology for behavioral economics. *The American Economic Review, 93*(5), 1449-1475. doi: 10.1257/000282803322655392

Kahneman, D. (2011). *Thinking, fast and slow*: Farrar, Straus and Giroux.

Kaymaz, N., Drukker, M., Lieb, R., Wittchen, H. U., Werbeloff, N., Weiser, M., . . . van Os, J. (2012). Do subthreshold psychotic experiences predict clinical outcomes in unselected non-help-seeking population-based samples? A systematic review and meta-analysis, enriched with new results. *Psychological Medicine, 42*(11), 2239-2253. doi: 10.1017/S0033291711002911

Krueger, F., & Grafman, J. (Eds.). (2013). *The neural basis of human belief systems*. New York, NY: Psychological Press.

Langdon, R. (2013). Folie a` deux and its lessons for two-factor theorists. *Mind & Language, 28*(1), 72-82. doi: 10.1111/mila.12009

Lawrie, S. M., Hall, J., McIntosh, A. M., Owens, D. G. C., & Johnstone, E. C. (2010). The 'continuum of psychosis': scientifically unproven and clinically impractical. *The British Journal of Psychiatry, 197*(6), 423-425. doi: 10.1192/bjp.bp.109.072827

Linscott, R. J., & van Os, J. (2013). An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: On the pathway from proneness to persistence to dimensional expression across mental disorders. *Psychological Medicine, 43*(6), 1133-1149. doi: 10.1017/S0033291712001626

Matthews, R. J. (2013). Belief and belief's penumbra. In N. Nottelman (Ed.), *New Essays on Belief* (pp. 100-123). New York, NY: Palgrave Macmillan.

McKay, R. T., & Dennett, D. C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences, 32*(6), 493-561. doi: 10.1017/S0140525X09990975

Mesoudi, A. (2004). Is human cultural evolution Darwinian? Evidence reviewed from the perspective of "The Origin of Species". *Evolution, 58*(1), 1-11. doi: 10.1111/j.0014-3820.2004.tb01568.x

Mesoudi, A. (2011). *Cultural evolution: how Darwinian theory can explain human culture and synthesize the social sciences*. Chicago, Il: University of Chicago Press.

Murphy, D. (2006). *Psychiatry in the scientific image*. Cambridge, MA: MIT Press.

Murphy, D. (2011). The folk epistemology of delusions. *Neuroethics, 5*(1), 19-22. doi: 10.1007/s12152-011-9125-5

Murphy, D. (2013). Delusions, modernist epistemology and irrational belief. *Mind & Language, 28*(1), 113-124. doi: 10.1111/mila.12012

Ngui, P. W. (1969). The Koro epidemic in Singapore. *Australian and New Zealand Journal of Psychiatry, 3*, 263-266.

Norenzayan, A. (2013). *Big gods: how religion transformed cooperation and conflict*. Princeton, NJ: Princeton University Press.

Norenzayan, A., Shariff, A., Willard, A. K., Slingerland, E., Gervais, W. M., McNamara, R. A., & Henrich, J. (in press). The cultural evolution of prosocial religions. *Behavioral and Brain Sciences, Advance online publication*. doi: 10.1017/s0140525x14001356

Peters, E. R., Joseph, S. A., Day, S., & Garety, P. A. (2004). Measuring delusional ideation: the 21-Item Peters et al. Delusions Inventory (PDI). *Schizophrenia Bulletin, 30*(4), 1005-1022. doi: 10.1093/oxfordjournals.schbul.a007116

Peters, E. R., Joseph, S. A., & Garety, P. A. (1999). Measurement of delusional ideation in the normal population: introducing the PDI (Peters et al. Delusions Inventory). *Schizophrenia Bulletin, 25*(3), 553-576. doi: 10.1093/oxfordjournals.schbul.a033401

Purzycki, B. G., & Sosis, R. (2009). The religious system as adaptive: cognitive flexibility, public displays, and acceptance. In E.V. & W. S. (Eds.), The biological evolution of religious mind and behavior (pp. 243-256). Berlin: Springer

Radden, J. (2013). Delusions redux. *Mind & Language, 28*(1), 125-139. doi: 10.1111/mila.12013

Richerson, P., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., . . . Zefferman, M. R. (in press). Cultural group selection plays an essential role in explaining human cooperation: a sketch of the evidence. *Behavioral and Brain Sciences, Advance online publication*. doi: 10.1017/s0140525x1400106x

Richerson, P., & Boyd, R. (2005). *Not by genes alone: how culture transformed human evolution*. Chicago, Il: University of Chicago Press.

Ross, R. M., & Atkinson, Q. D. (in press). Folktale transmission in the Arctic provides evidence for high bandwidth social learning among hunter-gatherer groups. *Evolution and Human Behavior.* doi: 10.1016/j.evolhumbehav.2015.08.001

Ross, R. M., & Atkinson, Q. D. (in press). Cultural differentiation does not entail group-level structure: the case for geographically explicit analysis. *Behavioral and Brain Sciences*.

Ross, R. M., Greenhill, S. J., & Atkinson, Q. D. (2013). Population structure and cultural geography of a folktale in Europe. *Proceedings of the Royal Society B-Biological Sciences, 280*(1756), 20123065. doi: 10.1098/rspb.2012.3065

Ross, R. M., McKay, R., Coltheart, M., & Langdon, R. (2015). Jumping to conclusions about the beads task? A meta-analysis of delusional ideation and data-gathering. *Schizophrenia Bulletin, Advance online publication*. doi: 10.1093/schbul/sbu187

Ross, R. M., Pennycook, G., Mckay, R., Gervais, W. M., Langdon, R., & Coltheart, M. (submitted). Analytic cognitive style, not delusional ideation, predicts data gathering in a large beads task study. *Manuscript submitted for publication*.

Sachdev, P. S. (1985). Koro epidemic in north-east India. *Australian and New Zealand Journal of Psychiatry, 19*, 433-438. doi: 10.1080/00048678509158852

Schwitzgebel, E. (2015). Belief. In E. N. Zaltam (Ed.), *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/belief/.

Sommer, I. E. (2010). The continuum hypothesis of psychosis: David's criticisms are timely. *Psychological Medicine, 40*(12), 1959-1961. doi: 10.1017/S0033291710000218

Speechley, W. J., Murray, C. B., McKay, R. M., Munz, M. T., & Ngan, E. T. C. (2010). A failure of conflict to modulate dual-stream processing may underlie the formation and maintenance of delusions. *European Psychiatry, 25*(2), 80-86. doi: 10.1016/j.eurpsy.2009.05.012

Speechley, W. J., & Ngan, E. T. C. (2008). Dual-stream modulation failure: a novel hypothesis for the formation and maintenance of delusions in schizophrenia. *Medical Hypotheses, 70*(6), 1210-1214. doi: 10.1016/j.mehy.2007.11.017

Speechley, W. J., Whitman, J. C., & Woodward, T. S. (2010). The contribution of hypersalience to the "jumping to conclusions" bias associated with delusions in schizophrenia. *Journal of Psychiatry & Neuroscience, 35*(1), 7-17. doi: 10.1503/jpn.090025

Speechley, W. J., Woodward, T. S., & Ngan, E. T. (2013). Failure of conflict to modulate central executive network activity associated with delusions in schizophrenia. *Frontiers in Psychiatry, 4*, e113. doi: 10.3389/fpsyt.2013.00113

Stafford, M. R., Jackson, H., Mayo-Wilson, E., Morrison, A. P., & Kendall, T. (2013). Early interventions to prevent psychosis: systematic review and meta-analysis. *BMJ, 346*, f185. doi: 10.1136/bmj.f185

Stanovich, K. E. (2011a). Normative models in psychology are here to stay. *Behavioral and Brain Sciences, 34*(5), 268-269. doi: 10.1017/s0140525x11000161

Stanovich, K. E. (2011c). *Rationality and the reflective mind*. Oxford, UK: Oxford University Press.

Stephens, G. L., & Graham, G. (2004). Reconceiving delusion. *International Review of Psychiatry, 16*(3), 236-241. doi: 10.1080/09540260400003982

Ventura, J., F., G. M., Shaner, A., & Lieberman, R. P. (1993). Training and quality assurance with the Brief Psychiatric Ratings Scale: "The Drift Buster". *International Journal of Methods in Psychiatric Research, 3*, 221-224.

West, S. A., El Mouden, C., & Gardner, A. (2011). Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior, 32*(4), 231-262. doi: 10.1016/j.evolhumbehav.2010.08.001

Whiten, A., Hinde, R. A., Laland, K. N., & Stringer, C. B. (2011). Culture evolves. *Philosophical Transactions of the Royal Society B-Biological Sciences, 366*(1567), 938-948. doi: 10.1098/rstb.2010.0372

Wilson, D. S. (2002). *Darwin's cathedral: evolution, religion, and the nature of society*. Chicago, Il: University of Chicago Press.

Chapter Two: Jumping to conclusions about the beads task? A meta-analysis of delusional ideation and data gathering

2.1 Abstract

It has been claimed that delusional and delusion-prone individuals have a tendency to gather less data before forming beliefs. Most of the evidence for this "jumping to conclusions" (JTC) bias comes from studies using the "beads task" data-gathering paradigm. However, the evidence for the JTC bias is mixed. We conducted a random-effects meta- analysis of individual participant data from 38 clinical and nonclinical samples (n = 2,237) to investigate the relationship between data gathering in the beads task (using the "draws to decision" measure) and delusional ideation (as indexed by the "Peters et al Delusions Inventory"; PDI). We found that delusional ideation is negatively associated with data gathering ($r_s = -0.10$, 95% *CI* [−0.17, −0.03]) and that there is heterogeneity in the estimated effect sizes (*Q*-stat $p = .03$, $I^2 = 33$). Subgroup analysis revealed that the negative association is present when considering the 23 samples (n = 1,754) from the large general population subgroup alone ($r_s = -0.10$, 95% *CI* [−0.18, −0.02]) but not when considering the eight samples (n = 262) from the small current delusions subgroup alone ($r_s = -0.12$, 95% *CI* [−0.31, 0.07]). These results provide some provisional support for continuum theories of psychosis and cognitive models that implicate the JTC bias in the formation and maintenance of delusions.

2.2 Introduction

In a now classic study, the "beads task" (Phillips & Edwards, 1966) was adapted to examine the relationship between delusions and data gathering (Huq, Garety, & Hemsley, 1988). Participants were shown two jars of beads, a mostly pink jar (85 pink beads; 15 green beads) and a mostly green jar (85 green beads; 15 pink beads). The jars were then hidden and participants were shown a sequence of beads apparently being drawn from one of the two jars (the sequence was actually prespecified by the experimenter). After each draw, participants were asked if they were ready to make a decision about which jar the beads were being drawn from or if they would like to see another bead. This study found that people with delusions made a decision about which jar the beads were being drawn from on the basis of significantly fewer beads than controls. This study has inspired a large empirical literature, and primarily on the basis of evidence from studies using the beads task paradigm it has been argued that people with delusions show a "jumping to conclusions" (JTC) bias: they are willing to accept hypotheses on the basis of less evidence than non-delusional people (Fine, Gardner, Craigie, & Gold, 2007; Garety & Freeman, 1999; Garety & Freeman, 2013).

It has long been argued that the positive symptoms of psychosis—delusions and hallucinations—lie at the extreme end of a continuum of similar subclinical phenomena in the general population (Bentall, 2003; Chapman & Chapman, 1980; Claridge, 1985; Meehl, 1962; Van Os, Hanssen, Bijl, & Ravelli, 2000). The existence of a "psychosis continuum" is supported by two recent meta-analyses (Linscott & van Os, 2013; van Os, Linscott, Myin-Germeys, Delespaul, & Krabbendam, 2009). Furthermore, it has been argued that the syndrome-based diagnostic categories of psychiatry impede progress in understanding the aetiology of mental illnesses and research should be reoriented to focus on vulnerability traits and symptoms across diagnostic categories and within the general population (Cuthbert, 2014; Hyman, 2010; Insel et al., 2010). Notably, the American National Institute of Mental Health (NIMH) has recently released a strategic plan that proposes abandoning syndrome-based classifications of the Diagnostics and Statistical Manual of Mental Disorders (DSM) to examine "the full range of variation, from normal to abnormal, among the fundamental components [dimensions] to improve understanding of what is typical versus pathological." (NIMH, 2008)

A dimensional approach that examines variation in the general population could provide insight into clinical delusions. In particular, evidence for an association between the JTC bias and delusional ideation in the general population would provide support for influential cognitive models that implicate the JTC bias in the aetiology of clinical delusions (Bentall, Corcoran, Howard, Blackwood, & Kinderman, 2001; Bentall, Fernyhough, Morrison, Lewis, & Corcoran, 2007; Freeman, 2007; Freeman & Garety, 2014; Garety, Bebbington, Fowler, Freeman, & Kuipers, 2007; Garety, Kuipers, Fowler, Freeman, & Bebbington, 2001; Morrison, 2001). The Peters *et al.* Delusions Inventory (PDI; Peters, Joseph, & Garety, 1999) is by far the most widely used measure of delusional ideation in the beads task literature. It is a self-report questionnaire that asks people if they have ever had particular delusion-like experiences. For example, one item asks, "Do you ever feel as if there is a conspiracy against you?" For each item endorsed people are asked to rate the degree of associated distress, preoccupation, and conviction separately on five-point Likert scales. The original PDI has 40 items (Peters, Joseph, et al., 1999), but a 21-item version with comparable psychometric properties is also widely used (Peters, Joseph, Day, & Garety, 2004). On average, patients diagnosed with delusions report higher PDI scores than healthy controls (Peters et al., 2004; Peters, Joseph, et al., 1999); on average, members of new religious movements report PDI scores that fall between those of delusional and general populations (Peters, Day, McKenna, & Orbach, 1999); and PDI scores correlate moderately strongly with observer-rated delusions using a structured clinical interview (Lincoln, Ziegler, Lullmann, Muller, & Rief, 2010). Such findings suggest that the PDI is a valid measure of thoughts that lie on a continuum with delusional beliefs.

Studies of the association between the JTC bias and delusions do not always provide consistent results, casting doubt on cognitive models implicating the JTC bias in delusion formation and maintenance. Consequently, careful evaluation of the overall weight of evidence is needed. A systematic review by Garety and Freeman (1999) argued that seven of eight beads task studies provided evidence that the JTC bias is associated with delusions. A follow-up by Garety and Freeman (2013) reviewed 53 new beads task studies (and eight studies using other probabilistic reasoning tasks) and concluded that "the clear majority of these studies ... confirmed that JTC is characteristic of people with delusions" (p. 328). Although this "vote counting" approach is commonly employed in systematic reviews, it is known to have significant limitations (Bushman & Wang, 2009; Schmidt & Hunter, 2014). First, it does not take into consideration features of studies (such as sample size) that ought to result in some studies being weighed more heavily than others. Second, it does not test or control for publication bias, selective reporting bias, or other biases that can inflate the evidence for an effect. Third, it does not quantify effect sizes.

The limitations of systematic reviews make meta-analysis a crucial tool for integrating evidence across studies (Cumming, 2012; Hedges & Olkin, 1985; Schmidt & Hunter, 2014). A 2007 meta-analysis by Fine *et al.* (2007) examined 12 studies. They found that one measure of the JTC bias reached statistical significance ("draws to decision"), but another three measures did not. Although useful, this meta- analysis has limitations. First, it used multiple effect sizes from the same samples (47 effect sizes from 22 samples) to estimate a single underlying construct (the JTC bias), which violates the assumption of statistical independence (Borenstein, Hedges, Higgins, & Rothstein, 2009; Schmidt & Hunter, 2014). Second, the Stouffer method of meta-analysis was used (Mosteller & Bush, 1954), which makes the problematic assumption that all studies are sampled from a single population (Schmidt & Hunter, 2014). Third, the possibility of publication bias and selective reporting bias was not examined. Fourth, effect sizes were not quantified.

Taylor, Hutton, and Dudley (2014) recently preregistered a rationale and protocol for a meta-analysis that will examine whether the JTC bias is associated with clinical delusions. This protocol is methodologically sophisticated and promises to address limitations of the earlier systematic reviews and meta-analysis. Nevertheless, because this protocol focuses on between- group differences it will not speak directly to the hypothesis that delusions lie at the extreme end of a continuum of subclinical phenomena within the general population (Bentall et al., 2001; Bentall et al., 2007; Freeman, 2007; Freeman & Garety, 2014; Garety et al., 2007; Garety et al., 2001; Morrison, 2001).

It has been argued that a variety of questionable research practices are prevalent in the social sciences, and an anonymous survey suggests that selective reporting of results is particularly widespread (John, Loewenstein, & Prelec, 2012). Selective reporting bias can result in false-positives well above the nominal rate of 5% (Simmons, Nelson, & Simonsohn, 2011), even when researchers strive to report their results dispassionately and honestly (Gelman & Loken, in press). Selective reporting bias can be curtailed by direct replication of experiments, ideally with preregistered protocols, and open access to raw data. However, due to the value placed on innovation in the social sciences, direct replication is extremely rare (Makel, Plucker, & Hegarty, 2012). In this respect, the beads task literature fits the typical profile of social science research. We were unable to identify any beads task study that directly replicated an earlier study, or used a preregistered protocol, or provided open access to raw data. Consequently there is considerable scope for selective reporting bias.

In the present meta-analysis, we tested the hypothesis that delusional ideation is negatively associated with data gathering in the general population and clinical populations. We did not use the effect sizes reported in publications. Instead we acquired raw data for each study that met our inclusion criteria and calculated the precise effect size of interest for each sample: the association between draws to decision on the beads task and PDI scores. This "individual participant data" approach offers numerous advantages over conventional meta-analysis and is considered to be the "gold standard" of systematic review (Cooper & Patall, 2009; Stewart & Tierney, 2002; Stewart, Tierney, & Clarke, 2008). Two advantages are particularly salient when considering the beads task literature. First, beads task studies typically report differences *between* samples *only*. That is, variation *within* delusional samples and *within* nondelusional samples, which is crucial for testing continuum models, is not always reported. Using individual participant data we were able to examine this crucial, but neglected, variation. Second, by consistently applying the same screening criteria and statistical tests to samples from different studies we were able to circumvent selective reporting bias and related biases.

2.3 Methods

2.3.1 Search strategy

We used two strategies for identifying studies for possible inclusion in our meta-analysis. First, we assessed for eligibility studies tabulated in the systematic review by Garety and Freeman (2013). They reported using three search techniques. First, they searched the Web of Science and PubMed databases using the following search terms: "jump to conclusions" and "delusions"; "jump to conclusions" and "schizophrenia"; "jump to conclusions" and "psychosis"; "jump to conclusions" and "paranoia". Second, they consulted five widely cited review articles on delusions (Bentall & Taylor, 2006; Dudley & Over, 2003; Fine et al., 2007; Freeman, 2007; Garety & Freeman, 1999). Third, they manually searched "early view" articles in the journals *Schizophrenia Bulletin*, *Schizophrenia Research*, *British Journal of Clinical Psychology*, *Behaviour Research and Therapy*, *Journal of Behavioural Research and Experimental Psychiatry*, *Psychological Medicine*, and *Journal of Abnormal Psychology*. In addition, we searched for articles published from 2011 to the present and "early view" articles using the same search techniques as Garety and Freeman (2013) to identify studies that might have been published after they completed their search.

Second, we used Google Scholar's cited reference search functionality to identify studies that had cited the article that introduced the 40-item PDI (Peters, Joseph, et al., 1999) or the 21-item PDI from 2011 to the present (Peters et al., 2004). Of the articles identified by Google Scholar, we inspected for possible inclusion those that had titles that indicated that they might include the beads task. Our literature search was completed July 10, 2014, and is inclusive of studies published up to that date.

2.3.2 Inclusion criteria

We categorised studies as eligible for inclusion in our meta-analysis if they met two inclusion criteria. First, the study used either a 40-item or 21-item PDI. We included all studies that used a PDI, even if the PDI had been modified. This meant that we included one study that used a version of the PDI that measured preoccupation only (Ochoa et al., 2014) and three studies that did not use the three PDI subscales but used only the initial "yes/no" question (Bentall et al., 2009; Langdon, Still, Connors, Ward, & Catts, 2013; Langdon, Ward, & Coltheart, 2010). Second, the study used a standard two jar draws to decision version of the beads task. We did not include studies that used "beads task-like" data-gathering paradigms (such as "emotional beads task" or "fishing task" paradigms) because we wanted tasks to be as directly comparable as possible.

We are interested in delusion ideation across syndrome-based diagnostic categories, so we did not exclude clinical groups that did not have a diagnosis of schizophrenia. We emailed the authors of all eligible studies with a request for raw data from their published and unpublished studies. We succeeded in sourcing raw data for all eligible published studies (bar one: Warman, 2008) and one currently unpublished study (R. Ephraums and R. P. Balzan, unpublished data). All studies we sourced that met our two inclusion criteria were included in the meta-analysis. See Table 2.1 for the full list of studies and their characteristics.

Table 2.1 Characteristics of samples included in the meta-analysis.

Sample	Subgroup	Participants	Mean Age	% Female	% Easy	Trials	Note
Bensi et al 2010	General Population	140	25.82	55	50	2	
Bentall et al 2009 (a)	General Population	63	56.32	63.49	0	3	1,2
Bentall et al 2009 (b)	Current Delusions	83	49.22	45.78	0	3	1,2
Bentall et al 2009 (c)	Previous Delusions	27	34.7	37.04	0	3	1,2
Bentall et al 2009 (d)	Anxiety or Depression	55	63.49	60	0	3	1,2
Broome et al 2007 (a)	General Population	22	24.87	N/A	50	2	1
Broome et al 2007 (b)	At Risk	27	24.56	N/A	50	2	1
Cafferkey et al 2013 (a)	General Population	133	22.86	68.42	100	1	
Cafferkey et al 2013 (b)	General Population	136	26.88	71.32	0	1	
Colbert & Peters 2002	General Population	68	41.21	55.88	100	1	
Ephraums & Balzan 2014	General Population	99	23.38	75.76	50	2	
Jacobsen et al 2012 (a)	General Population	16	34.19	56.25	50	2	
Jacobsen et al 2012 (b)	Current Delusions	16	39.5	43.75	50	2	
Jacobsen et al 2012 (c)	OCD	32	35.66	62.5	50	2	
Keefe & Warman 2011	General Population	132	21.42	78.79	0	4	
Langdon et al 2010 (a)	General Population	34	32.03	23.53	100	1	2
Langdon et al 2010 (b)	Current Delusions	29	35.1	34.48	100	1	2
Langdon et al 2013 (a)	General Population	19	20.79	10.53	100	1	2
Langdon et al 2013 (b)	Current Delusions	17	20.59	0	100	1	2
Lim et al 2012 (a)	General Population	63	23.95	74.6	50	2	
Lim et al 2012 (b)	Current Delusions	25	24.6	36	50	2	
Lim et al 2012 (c)	NRM	32	31.03	37.5	50	2	
Lincoln et al 2010 (a)	General Population	68	33.76	38.24	50	6	
Lincoln et al 2010 (b)	Current Delusions	44	35.48	31.82	50	6	
Lincoln et al 2010 (c)	Previous Delusions	27	30.59	29.63	50	6	
McKay et al 2006	General Population	57	20.96	63.16	100	1	
Menon et al 2013	General Population	121	31.05	64.46	0	1	
Ochoa et al 2014	General Population	57	45.07	40.35	50	2	3
Peters & Garety 2006 (a)	General Population	36	27.72	50	100	1	
Peters & Garety 2006 (b)	Current Delusions	18	32.22	11.11	100	1	
Peters & Garety 2006 (c)	Anxiety or Depression	21	41.19	47.62	100	1	
Rodier et al 2011	General Population	78	29.24	57.69	100	1	
So et al 2008 (a)	General Population	30	20.07	66.67	50	2	
So et al 2008 (b)	Current Delusions	30	21.6	56.67	50	2	
Warman & Martin 2006	General Population	199	21.11	77.39	100	4	
Warman et al 2007	General Population	59	21.39	74.58	0	4	
White & Mansell 2009	General Population	39	19.44	84.62	50	2	
Ziegler et al 2008	General Population	85	24.31	58.82	100	3	

Note: OCD, obsessive-compulsive disorder; NRM, new religious movement. Note 1 = 40-item Peters *et al*. Delusions Inventory (PDI) for total PDI score; Note 2 = initial "yes/no" question for total PDI score; Note 3 = preoccupation for total PDI score.

2.3.3 Data coding

When possible, we calculated total PDI scores for each participant by adding the number of "yes" responses to scores from the three subscales. This was possible for 27 samples (see Table 2.1 for exceptions). Twenty-five samples were tested using the 21-item PDI (Cafferkey, Murphy, & Shevlin, 2013; Colbert & Peters, 2002; Jacobsen, Freeman, & Salkovskis, 2012; Langdon et al., 2013; Langdon et al., 2010; Lim, Gleeson, & Jackson, 2012; McKay, Langdon, & Coltheart, 2006; Menon et al., 2013; Ochoa et al., 2014; Peters & Garety, 2006; Rodier et al., 2011; So, Freeman, & Garety, 2008; White & Mansell, 2009; R. Ephraums and R. P. Balzan, unpublished data) and 13 using the 40-item PDI (Bentall et al., 2009; Broome et al., 2007; Keefe & Warman, 2011; Lincoln, Ziegler, Mehl, & Rief, 2010; Warman, Lysaker, Martin, Davis, & Haudenschield, 2007; Warman & Martin, 2006; Ziegler, Rief, Werner, Mehl, & Lincoln, 2008). When calculating total PDI scores for samples that used the 40-item PDI, we included only those 21 items that appear in the 21-item PDI. Raw data obtained for 6 samples that used the 40-item PDI did not include scores for individual items (Bentall et al., 2009; Broome et al., 2007); in these cases we used the total PDI score for all 40 items. Eight samples used a version of the PDI that did not include the subscales (Bentall et al., 2009; Langdon et al., 2013; Langdon et al., 2010); in these cases we used "yes" responses to the initial questions to calculate the PDI total score. For the sample that used a version of the PDI that measured preoccupation only (Ochoa et al., 2014), we used the preoccupation score to calculate the PDI total score.

Samples varied with respect to the maximum number of beads participants were able to request before making a decision about which jar the beads were being drawn from. Participants were not told what this limit was. If they reached this limit they were asked to make a decision. Following standard practice, we retained data from participants who reached the limit. Samples also varied with respect to the number of beads task trials presented to each participant. For analysis, we calculated a mean draws to decision score across trials for each participant. We coded beads task trials with a ratio of beads of 60:40 as "difficult" and beads task trials with ratios of beads of either 85:15 or 80:20 as "easy," which we used to calculate the percentage of trials that were "easy" for each sample.

Inspection of raw data occasionally revealed instances of typographic errors. When we identified such errors we attempted to infer correct values. When this was not possible we recoded erroneous values as missing data. Participants who had missing data for PDI score, draws to decision, age, or gender were removed prior to analysis. The single exception was the study by Broome *et al.* (2007) that did not code for gender; we included participants from this study in all analyses that did not include gender as a variable. In total, 58 participants (2.5% of participants) were removed due to missing data.

2.3.4 Statistical analysis

Inspection of PDI scores and mean draws to decision revealed that neither variable was normally distributed, so we used the nonparametric Spearman's rank correlation coefficient (r_s) for analysis. As per the method advocated by Hedges and Olkin (1985), we converted r_s scores to their associated Fisher's *z*-scores for estimating uncertainty in effect sizes and back-transformed Fisher's *z*-scores to r_s scores for interpretation.

Meta-analysis was conducted using the software OpenMEE (Dietz et al., 2014; Wallace et al., 2012) the R package Metafor (Team, 2013; Viechtbauer, 2010). We used a random effects model (Borenstein et al., 2009; Cumming, 2012) and examined heterogeneity in estimates of r_s for the overall group of samples and diagnostic subgroups using the Q statistic (Cochran, 1954) and the I^2 index (Higgins, Thompson, Deeks, & Altman, 2003). The Q statistic can be underpowered (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006), so we paid some attention to I^2 indices even in cases of nonsignificant Q statistics. We used the Sidik–Jonkman method for estimating heterogeneity because it provides more accurate estimates than the more widely used DerSimonian–Laird method (IntHout, Ioannidis, & Borm, 2014). To examine potential moderators of effect sizes, we used random-effects meta-regression (Thompson & Higgins, 2002) with a separate meta-regression for each potential moderator. We assessed the evidence for publication bias using a funnel plot and Egger's regression test for funnel plot asymmetry (Egger, Smith, Schneider, & Minder, 1997; Sterne et al., 2011).

2.4 Results

2.4.1 Meta-analysis

Figure 2.1 shows a forest plot for the random-effects meta-analysis. The analysis indicates that there is a negative association between draws to decision and PDI score ($r_s = -0.10$, 95% *CI* [−0.17, −0.03], $n = 2{,}237$, $k = 38$; see the dark gray diamond in Figure 2.1). We found moderate heterogeneity (*Q*-stat $p = .03$, $I^2 = 33$), which suggests that the precise magnitude of the overall effect size should be interpreted with some degree of caution. Given that three samples with six trials (all of which came from the same study) might have a large influence on analyses (see section 2.4.3. below), we tested the robustness of the meta-analysis by re-running the meta-analysis with these three samples removed. We found that the negative association between draws to decision and PDI scores remains essentially the same ($r_s = -0.11$, 95% *CI* [−0.16, −0.06], $n = 2{,}098$, $k = 35$.

2.4.2 Subgroup analysis

We found a negative association between draws to decision and PDI when considering the general population subgroup alone ($r_s = -0.10$, 95% *CI* [−0.18, −0.02], $n = 1{,}754$, $k = 23$), but not in the current delusions subgroup alone ($r_s = -0.12$, 95% *CI* = −0.31, 0.07, $n = 262$, $k = 8$), the previous delusions subgroup alone ($r_s = 0.05$, 95% *CI* [−0.53, 0.63], $n = 54$, $k = 2$), or the anxiety or depression subgroup alone ($r_s = -0.04$, 95% CI [−0.28, 0.19], $n = 76$, $k = 2$; see respective light gray diamonds in Figure 2.1).

We found moderate heterogeneity within the general population subgroup (*Q*-stat $p = .03$, $I^2 = 40$) and substantial heterogeneity within the previous delusions subgroup (*Q*-stat $p = .03$, $I^2 = 78$), but we did not find statistically significant heterogeneity in the current delusions subgroup (Q-stat p = .11, I^2 = 40) or the anxiety or depression subgroup (Q-stat p = .78, I^2 = 0).

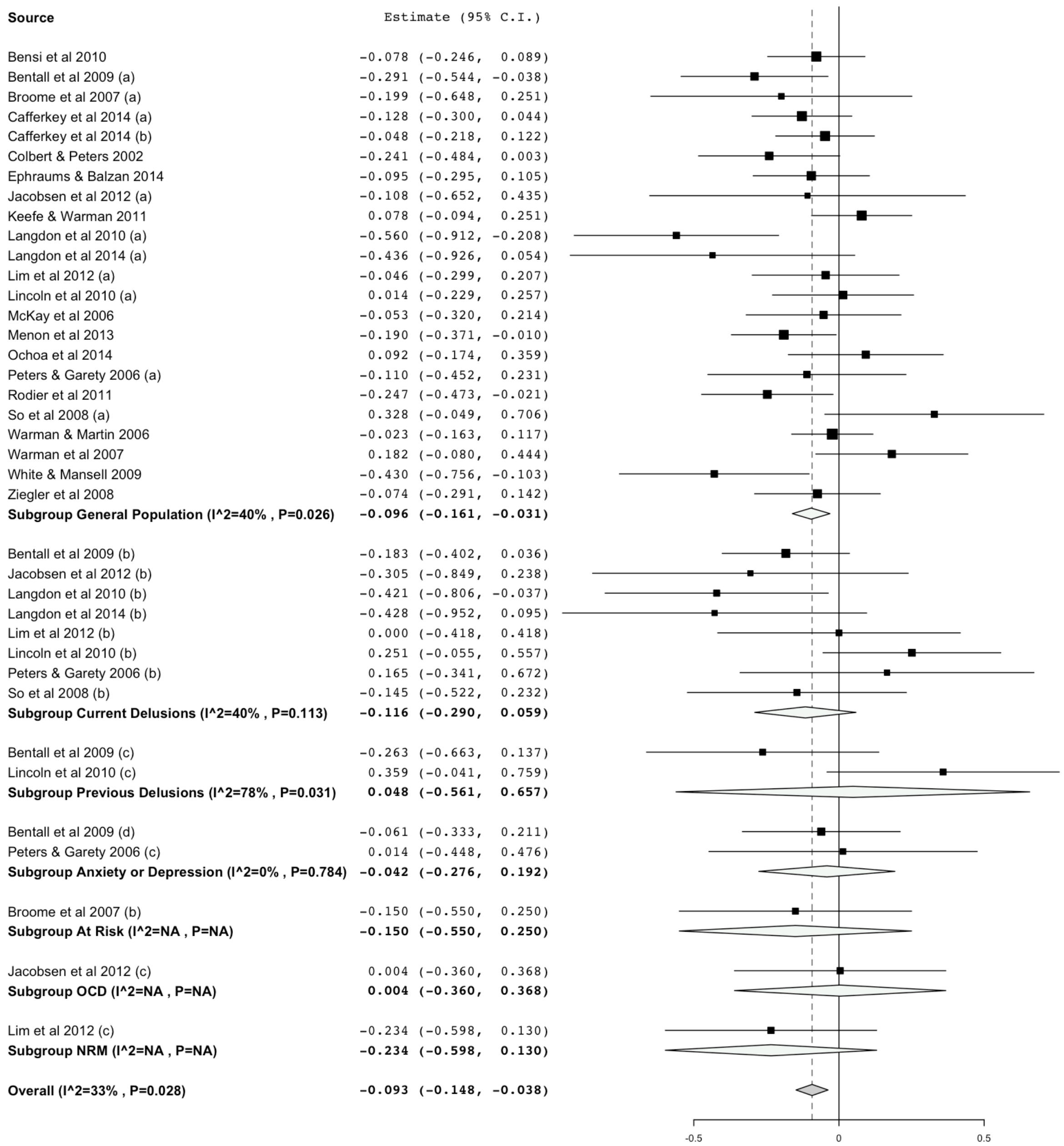


Figure 2.1 Forest plot of random effects meta-analysis showing effect sizes (r_s) for the association between draws to decision and Peters *et al*. Delusions Inventory. The black squares show effect sizes for each sample and are drawn proportional to the relative weighting of each sample in the analysis. The error bars show the 95% CI for each sample. The dark gray diamond shows the overall 95% CI. The light gray diamonds show the 95% CI

for each subgroup. The broken line shows the overall mean effect size estimate.

2.4.3 Meta-regression

Because there is evidence for heterogeneity in effect sizes we performed exploratory moderator analysis. Figure 2.2 shows that the number of trials is a statistically significant predictor of r_s ($\beta = 0.07$, $SE = 0.02$; 95% CI [0.02, 0.11]; $z = 3.02$, $p < .01$; $\alpha = -0.26$, $SE = 0.06$; 95% CI [−0.38, −0.13]; $z = -4.09$, $p < .01$). Visual inspection of Figure 2.2 suggests that the three samples with six trials (all of which came from the same study) might have a large influence on the regression. To test the robustness of this association we re-ran the meta-regression with these three samples removed and found that the number of trials is no longer a statistically significant predictor of r_s ($\beta = 0.06$, $SE = 0.03$; 95% CI [0.00, 0.13]; $z = 1.92$, $p = .06$; $\alpha = -0.25$, $SE = 0.08$; 95% CI [−0.40, −0.10]; $z = -3.27$, $p < .01$), which suggests that some caution is warranted when interpreting this association.

Other potential sample-level moderators were not found to be statistically significant predictors of effect size: mean age of participants ($\beta = 0.00$, $SE = 0.00$; 95% CI [−0.01, 0.01]; $z = -0.59$, $p = .56$; $\alpha = -0.04$, $SE = 0.11$; 95% CI [−0.25, 0.18]; $z = -0.32$, $p = .75$); percentage of females (male = 0, female = 1; $\beta = 0.00$, $SE = 0.00$; 95% CI [0.00, 0.01]; $z = 0.84$, $p = .40$; $\alpha = -0.19$, $SE = 0.12$; 95% CI [−0.42, 0.04]; $z = -1.60$, $p = .11$); and percentage of trials using an easy version of the beads task ($\beta = 0.00$, $SE = 0.00$; 95% CI [−0.00, 0.00]; $z = -1.07$, $p = .28$; $\alpha = -0.04$, $SE = 0.06$; 95% CI [−0.16, 0.08]; $z = -0.71$, $p = .48$).

2.4.4 Publication bias

Figure 2.3 shows a funnel plot. A negative effect size is predicted, so publication bias would be expected to manifest as a gap in the bottom right region of the funnel plot. There is no obvious gap here or elsewhere. In addition, the plot appears to be relatively symmetric and Egger's regression test for funnel plot asymmetry was nonsignificant ($z = -0.74$, $p = .46$). Overall, we found no evidence for publication bias. Nevertheless, absence of evidence should be interpreted with some caution, because even when publication bias is present it can be difficult to identify (Sterne et al., 2011; Terrin, Schmid, & Lau, 2005).

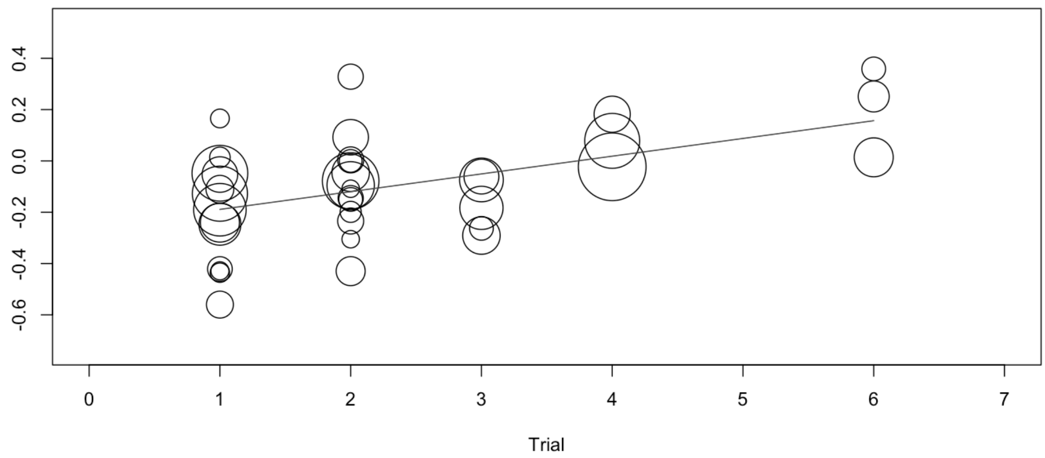


Figure 2.2 Regression plot showing the random-effects meta-regression examining the relationship between number of beads task trials and effect sizes. Effect sizes (r_s) are plotted as circles with the size of the circles drawn proportional to the relative weight of the sample in the analysis.

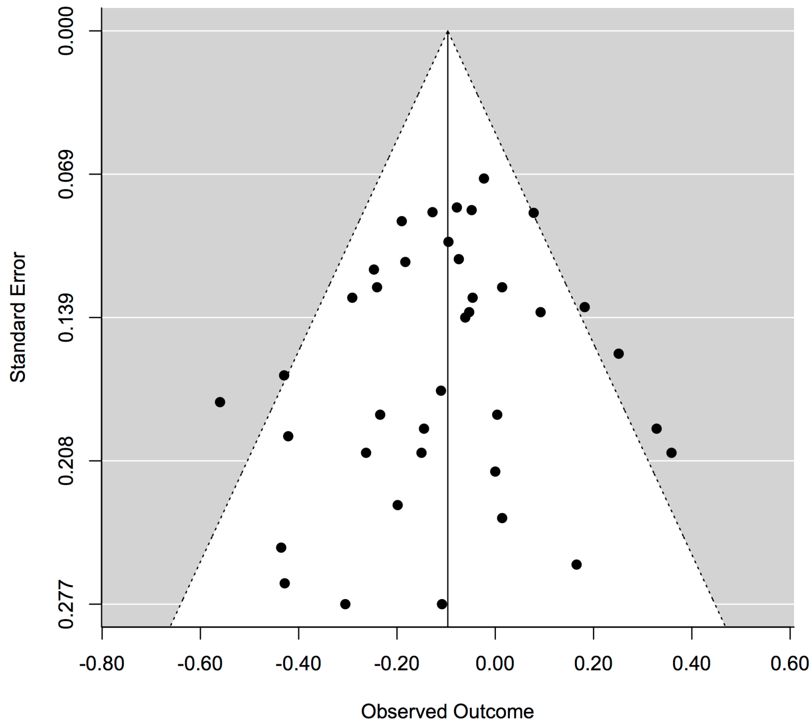


Figure 2.3 Funnel plot showing effect size (r_s) plotted against standard error.

2.5 Discussion

2.5.1 Summary of findings

Overall, these results suggest that people with higher PDI scores tend to request fewer beads before making a decision about which jar beads are being drawn from than people with lower PDI scores. The overall effect size is small, and exhibits moderate heterogeneity. This negative association between PDI scores and draws to decision is present when considering the general population subgroup alone, but not when considering the current delusions subgroup alone. We found no evidence to suggest that effect size estimates have been inflated as a result of publication bias.

2.5.2 Discussion of findings

Overall, these results provide some provisional support for continuum theories of psychosis (Bentall, 2003; Chapman & Chapman, 1980; Claridge, 1985; Linscott & van Os, 2013; Meehl, 1962; Van Os et al., 2000; van Os et al., 2009) and cognitive models that implicate the JTC bias in the formation and maintenance of delusions (Bentall et al., 2001; Bentall et al., 2007; Freeman, 2007; Freeman & Garety, 2014; Garety et al., 2007; Garety et al., 2001; Morrison, 2001). That we did not find evidence for an association in the current delusions subgroup could be considered to be of concern, as cognitive theories aim to account for clinical delusions not mere delusional ideation in the general population. Nonetheless, as Figure 2.1 shows, the confidence intervals for the general population subgroup and the current delusions subgroup overlap substantially and have almost identical means (in fact, the mean for the current delusions subgroup is slightly more negative than the mean for the general population subgroup). We suggest that the low statistical power of subgroup analysis (Schmidt & Hunter, 2014) is a more plausible interpretation of this counter-intuitive result than proposing that the association is present in the large general population subgroup (n = 1,754; k = 23) but not the small current delusions subgroup (n = 262; k = 8).

Although these results provide evidence for an association between delusional ideation and data gathering, it is important to consider the small effect sizes. One possible interpretation of the small effect sizes is that the JTC bias plays only a minor role in delusion formation and maintenance, despite the many published studies reporting that the JTC bias is more common in those with current delusions than controls (Garety & Freeman, 1999; Garety & Freeman, 2013). It may even be the case that the JTC bias would explain *no* additional variation in delusional ideation if other important dimensions of individual variation were taken into account. Such a possibility is consistent with one of the larger beads task studies included in our meta-analysis, which found no evidence for an association between paranoia (in the context of paranoid delusions) and jumping to conclusions once the association between paranoia and general cognitive performance had been controlled for (Bentall et al., 2009). We anticipate that the upcoming meta-analysis by Taylor *et al.* (2014) will help clarify because they aim to quantify differences in data gathering between groups with delusions and control groups while exploring a variety of potential moderating variables.

Another possible interpretation of these small effect sizes is that widely used versions of the beads task paradigm might not be well suited to examining the JTC bias. This possibility is consistent with claims that studies that have used the beads task may have significant methodological limitations. First, if a participant asks to see only a small number of beads it is not always clear that they are *jumping* to conclusions on the basis of *insufficient* evidence (Maher, 2004). In many instances, very few beads need to be drawn for the posterior probability of one of the jars to be very high (e.g., when the ratio of colours is 85:15 and the first two beads are the same colour, the posterior probability of the corresponding jar is 0.97). Second, evidence from a "graded estimates" variant of the beads task suggests that people with delusions misunderstand task instructions more often than controls (Balzan, Delfabbro, Galletly, & Woodward, 2012). Third, the beads task is rarely incentivised and motivation might explain some differences in performance (van der Leer, Hartig, Goldmanis, & McKay, in press; van der Leer & McKay, 2013).

2.5.3 Limitations of the present study and directions for future research

The present study has two important limitations that are worth highlighting. First, there is variation in beads task protocols that is not examined. For instance, some beads tasks studies include a memory aid (the colours of the previously shown beads are presented) while others do not, and there is some evidence that the association between the JTC bias and schizophrenia might be related to memory demands (Menon, Pomarol-Clotet, McKenna, & McCarthy, 2006). Future research could examine whether other sources of variation moderate effect sizes. Second, the PDI was the only measure of delusional ideation that was used. This meant that only a limited number of studies that focused on schizophrenia patients were included. Future research could examine clinical scales that measure symptom severity in schizophrenia, such as the widely used Positive and Negative Syndrome Scale (PANNS; Kay, Fiszbein, & Opler, 1987).

We suggest that further progress in determining whether people with delusions jump to conclusions could be made by using meta-analysis to investigate between-group differences in beads task performance (Taylor et al., 2014), exploring new data-gathering paradigms that might overcome methodological limitations of the beads task (Sanford, Veckenstedt, Moritz, Balzan, & Woodward, 2014; Speechley, Whitman, & Woodward, 2010; van der Leer et al., in press; van der Leer & McKay, 2013; Whitman, Menon, Kuo, & Woodward, 2013; Woodward, Moritz, Cuttler, & Whitman, 2006; Woodward, Munz, LeClerc, & Lecomte, 2009), examining more closely what evidence is available to participants at the point when they decide to stop drawing beads (Jolley et al., 2014), controlling for important aspects of individual variation (Bentall et al., 2009), and revisiting the original—and methodologically sophisticated— beads task paradigm (Phillips & Edwards, 1966).

2.6 References

Balzan, R. P., Delfabbro, P. H., Galletly, C. A., & Woodward, T. S. (2012). Over-adjustment or miscomprehension? A re-examination of the jumping to conclusions bias. *Australian & New Zealand Journal of Psychiatry, 46*(6), 532-540. doi: 10.1177/0004867411435291

Bentall, R. P. (2003). *Madness explained: psychosis and human nature*. London, UK: Penguin.

Bentall, R. P., Corcoran, R., Howard, R., Blackwood, N., & Kinderman, P. (2001). Persecutory delusions: a review and theoretical integration. *Clinical Psychology Review, 21*(8), 1143-1192. doi: 10.1016/s0272-7358(01)00106-4

Bentall, R. P., Fernyhough, C., Morrison, A. P., Lewis, S., & Corcoran, R. (2007). Prospects for a cognitive-developmental account of psychotic experiences. *British Journal of Clinical Psychology, 46*, 155-173. doi: 10.1348/014466506x123011

Bentall, R. P., Rowse, G., Shryane, N., Kinderman, P., Howard, R., Blackwood, N., . . . Corcoran, R. (2009). The cognitive and affective structure of paranoid delusions: A transdiagnostic investigation of patients with Schizophrenia spectrum disorders and depression. *Archives of General Psychiatry, 66*(3), 236-247. doi: 10.1001/archgenpsychiatry.2009.1

Bentall, R. P., & Taylor, J. L. (2006). Psychological processes and paranoia: implications for forensic behavioural science. *Behavioral Science & the Law, 24*(3), 277-294. doi: 10.1002/bsl.718

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley & Sons, Ltd.

Broome, M. R., Johns, L. C., Valli, I., Woolley, J. B., Tabraham, P., Brett, C., . . . McGuire, P. K. (2007). Delusion formation and reasoning biases in those at clinical high risk for psychosis. *British Journal of Psychiatry, 51*, s38-42. doi: 10.1192/bjp.191.51.s38

Bushman, B. J., & Wang, M. D. (2009). Vote counting procedures in meta-analysis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 207-220). New York: Russell Sage.

Cafferkey, K., Murphy, J., & Shevlin, M. (2014). Jumping to conclusions: The association between delusional ideation and reasoning biases in a healthy student population. *Psychosis, 6*(3), 206-214. 1-9. doi: 10.1080/17522439.2013.850734

Chapman, L. J., & Chapman, J. P. (1980). Scales for rating psychotic and psychotic-like experiences as continua. *Schizophrenia Bulletin, 6*(3), 476-489. doi: 10.1093/schbul/6.3.476

Claridge, G. (1985). *Origins of mental illness: temperament, deviance and disorder*. Oxford, UK: Blackwell.

Cochran, W. (1954). The combination of estimates from different experiments. *Biometrics, 10*(1), 101-129. doi: 10.2307/3001666

Colbert, S. M., & Peters, E. R. (2002). Need for closure and jumping-to-conclusions in delusion-prone individuals. *Journal of Nervous and Mental Disease, 190*(1), 27-31. doi: 10.1097/00005053-200201000-00007

Cooper, H., & Patall, E. A. (2009). The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychological Methods, 14*(2), 165-176. doi: 10.1037/a0015565

Cumming, G. (2012). *Understanding the new statistics: effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Cuthbert, B. N. (2014). The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry, 13*(1), 28-35. doi: 10.1002/wps.20087

Dietz, G., Dahabreh, I. J., Gurevitch, J., Lajeunesse, M. J., Schmid, C. H., Trikalinos, T. A., & Wallace, B. C. (2014). OpenMEE: software for ecological and evolutionary meta-analysis. Available at (http://www.cebm.brown.edu/open_mee). (Accessed 30 July 2014).

Dudley, R. E. J., & Over, D. E. (2003). People with delusions jump to conclusions: A theoretical account of research findings on the reasoning of people with delusions. *Clinical Psychology & Psychotherapy, 10*(5), 263-274. doi: 10.1002/cpp.376

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*, 629-634. doi: 10.1136/bmj.315.7109.629

Fine, C., Gardner, M., Craigie, J., & Gold, I. (2007). Hopping, skipping or jumping to conclusions? Clarifying the role of the JTC bias in delusions. *Cognitive Neuropsychiatry, 12*(1), 46-77. doi: 10.1080/13546800600750597

Freeman, D. (2007). Suspicious minds: the psychology of persecutory delusions. *Clinical Psychology Review, 27*(4), 425-457. doi: 10.1016/j.cpr.2006.10.004

Freeman, D., & Garety, P. (2014). Advances in understanding and treating persecutory delusions: a review. *Social Psychiatry and Psychiatric Epidemiology, 49*(8), 1179-1189. doi: 10.1007/s00127-014-0928-7

Garety, P. A., Bebbington, P., Fowler, D., Freeman, D., & Kuipers, E. (2007). Implications for neurobiological research of cognitive models of psychosis: a theoretical paper. *Psychological Medicine, 37*(10), 1377-1391. doi: 10.1017/s003329170700013x

Garety, P. A., & Freeman, D. (1999). Cognitive approaches to delusions: a critical review of theories and evidence. *British Journal of Clinical Psychology, 38* (2), 113-154. doi: 10.1348/014466599162700

Garety, P. A., & Freeman, D. (2013). The past and future of delusions research: from the inexplicable to the treatable. *British Journal of Psychiatry, 203*(5), 327-333. doi: 10.1192/bjp.bp.113.126953

Garety, P. A., Joyce, E., Jolley, S., Emsley, R., Waller, H., Kuipers, E., . . . Freeman, D. (2013). Neuropsychological functioning and jumping to conclusions in delusions. *Schizophrenia Research, 150*(2-3), 570-574. doi: 10.1016/j.schres.2013.08.035

Garety, P. A., Kuipers, E., Fowler, D., Freeman, D., & Bebbington, P. E. (2001). A cognitive model of the positive symptoms of psychosis. *Psychological Medicine, 31*, 189-195. doi: 10.1017/s0033291701003312

Gelman, A., & Loken, E. (in press). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *American Scientist*.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analysis. *BMJ, 327*, 557-560. doi: 10.1136/bmj.327.7414.557

Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods, 11*(2), 193-206. doi: 10.1037/1082-989X.11.2.193

Huq, S. F., Garety, P. A., & Hemsley, D. R. (1988). Probabilistic judgments in deluded and non-deluded subjects. *Quarterly Journal of Experimental Psychology Section A, 40*(4), 801-812. doi: 10.1080/14640748808402300

Hyman, S. E. (2010). The diagnosis of mental disorders: The problem of reification. *Annual Review of Clinical Psychology, 6*, 155-179. doi: 10.1146/annurev.clinpsy.3.022806.091532

Insel, T. R., Cuthbert, B. N., Garvey, M., Heinssen, Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry, 167*(7), 748-751. doi: 10.1176/appi.ajp.2010.09091379

IntHout, J., Ioannidis, J., & Borm, G. F. (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology, 14*(25), 1471-2288. doi: 10.1186/1471-2288-14-25

Jacobsen, P., Freeman, D., & Salkovskis, P. (2012). Reasoning bias and belief conviction in obsessive-compulsive disorder and delusions: jumping to conclusions across disorders? *British Journal of Clinical Psychology, 51*(1), 84-99. doi: 10.1111/j.2044-8260.2011.02014.x

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*(5), 524-532. doi: 10.1177/0956797611430953

Jolley, S., Thompson, C., Hurley, J., Medin, E., Butler, L., Bebbington, P., . . . Garety, P. (2014). Jumping to the wrong conclusions? An investigation of the mechanisms of reasoning errors in delusions. *Psychiatry Research, 219*(2)*, 275-282*. doi: 10.1016/j.psychres.2014.05.051

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizoprhenia. *Schizophrenia Bulletin, 13*(2), 261-276. doi: 10.1093/schbul/13.2.261

Keefe, K. M., & Warman, D. M. (2011). Reasoning, delusion proneness and stress: an experimental investigation. *Clinical Psychology & Psychotherapy, 18*(2), 138-147. doi: 10.1002/cpp.683

Langdon, R., Still, M., Connors, M. H., Ward, P. B., & Catts, S. V. (2014). Jumping to delusions in early psychosis. *Cognitive Neuropsychiatry, 19*(3), 241-256. doi: 10.1080/13546805.2013.854198

Langdon, R., Ward, P. B., & Coltheart, M. (2010). Reasoning anomalies associated with delusions in Schizophrenia. *Schizophrenia Bulletin, 36*(2), 321-330. doi: 10.1093/schbul/sbn069

Lim, M. H., Gleeson, J. F., & Jackson, H. J. (2012). The jumping-to-conclusions bias in new religious movements. *The Journal of Nervous and Mental Disease, 200*(10), 868-875. doi: 10.1097/NMD.0b013e31826b6eb4

Lincoln, T. M., Ziegler, M., Lullmann, E., Muller, M. J., & Rief, W. (2010). Can delusions be self-assessed? Concordance between self- and observer-rated delusions in schizophrenia. *Psychiatry Research, 178*(2), 249-254. doi: 10.1016/j.psychres.2009.04.019

Lincoln, T. M., Ziegler, M., Mehl, S., & Rief, W. (2010). The jumping to conclusions bias in delusions: specificity and changeability. *Journal of Abnormal Psychology, 119*(1), 40-49. doi: 10.1037/a0018118

Linscott, R. J., & van Os, J. (2013). An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: On the pathway from proneness to persistence to dimensional expression across mental disorders. *Psychological Medicine, 43*(6), 1133-1149. doi: 10.1017/S0033291712001626

Maher, B. A. (2004). Delusion. In Adams, H. E. & Sutker, P.B. (Eds.) *Comprehensive handbook of psychopathology* (3rd ed.). New York, NY: Springer.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives in Psychological Science, 7*(6), 537-542. doi: 10.1177/1745691612460688

McKay, R., Langdon, R., & Coltheart, M. (2006). Need for closure, jumping to conclusions, and decisiveness in delusion-prone individuals. *The Journal of Nervous and Mental Disease, 194*(6), 422-426. doi: 10.1097/01.nmd.0000221353.44132.25

Meehl, P. E. (1962). Schizotaxia, schizotypy, schizophrenia. *American Psychologist, 17*, 827-838. doi: 10.1037/h0041029

Menon, M., Pomarol-Clotet, E., McKenna, P. J., & McCarthy, R. A. (2006). Probabilistic reasoning in schizophrenia: a comparision of performance of deluded and nondeluded schizophrenic patients and exploration of possibile cognitive underpinnings. *Cognitive Neuropsychiatry, 11*(6), 521-536. doi: 10.1080/13546800544000046

Menon, M., Quilty, L. C., Zawadzki, J. A., Woodward, T. S., Sokolowski, H. M., Boon, H. S., & Wong, A. H. C. (2013). The role of cognitive biases and personality variables in subclinical delusional ideation. *Cognitive Neuropsychiatry, 18*(3), 208-218. doi: 10.1080/13546805.2012.692873

Morrison, A. P. (2001). The interpretation of intrusions in psychosis: An integrative cognitive approach to hallucinations and delusions. *Behavioural and Cognitive Psychotherapy, 29*, 257-276. doi: 10.1017/s1352465801003010

Mosteller, F. M., & Bush, R. R. (1954). Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology: vol I. theory and method*. Cambridge, MA: Addison-Wesley.

NIMH. (2008). The National Institute of Mental Health Strategic Plan. from http://www.nimh.nih.gov/about/strategic-planning-reports/index.shtml - strategic-objective1

Ochoa, S., Haro, J. M., Huerta-Ramos, E., Cuevas-Esteban, J., Stephan-Otto, C., Usall, J., . . . Brebion, G. (2014). Relation between jumping to conclusions and cognitive functioning in people with schizophrenia in contrast with healthy participants. *Schizophrenia Research, 159*, 211-217. doi: 10.1016/j.schres.2014.07.026

Peters, E. R., Day, S., McKenna, J., & Orbach, G. (1999). Delusional ideation in religious and psychotic populations. *British Journal of Clinical Psychology, 38*, 83-96. doi: 10.1348/014466599162683

Peters, E. R., & Garety, P. (2006). Cognitive functioning in delusions: A longitudinal analysis. *Behavioural Research and Therapy, 44*(4), 481-514. doi: 10.1016/j.brat.2005.03.008

Peters, E. R., Joseph, S. A., Day, S., & Garety, P. A. (2004). Measuring Delusional Ideation: The 21-Item Peters et al. Delusions Inventory (PDI) *Schizophrenia Bulletin, 30*(4), 1005-1022. doi: 10.1093/oxfordjournals.schbul.a007116

Peters, E. R., Joseph, S. A., & Garety, P. A. (1999). Measurement of delusional ideation in the normal population: Introducing the PDI (Peters et al. Delusions Inventory). *Schizophrenia Bulletin, 25*(3), 553-576. doi: 10.1093/oxfordjournals.schbul.a033401

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology, 72*(3), 346-354. doi: 10.1037/h0023653

Rodier, M., Prevost, M., Renoult, L., Lionnet, C., Kwann, Y., Dionne-Dostie, E., . . . Debruille, J. B. (2011). Healthy people with delusional ideation change their mind with conviction. *Psychiatry Research, 189*(3), 433-439. doi: 10.1016/j.psychres.2011.06.018

Sanford, N., Veckenstedt, R., Moritz, S., Balzan, R. P., & Woodward, T. S. (2014). Impaired integration of disambiguating evidence in delusional schizophrenia patients. *Psychological Medicine, 44*(13), 2729-2738. 1-10. doi: 10.1017/s0033291714000397

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: correction error and bias in research findings* (3 ed.). Thousand Oaks, CA: Sage.

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. doi: 10.1177/0956797611417632

So, S. H., Freeman, D., & Garety, P. (2008). Impact of state anxiety on the jumping to conclusions delusion bias. *Australian & New Zealand Journal of Psychiatry, 42*(10), 879-886. doi: 10.1080/00048670802345466

Speechley, W. J., Whitman, J. C., & Woodward, T. S. (2010). The contribution of hypersalience to the “jumping to conclusions” bias associated with delusions in schizophrenia. *Journal of Psychiatry & Neuroscience, 35*(1), 7-17. doi: 10.1503/jpn.090025

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., . . . Higgins, J. P. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ, 343*, d4002. doi: 10.1136/bmj.d4002

Stewart, L. A., & Tierney, J. F. (2002). To IPD or not to IPD?: Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation & the Health Professions, 25*(1), 76-97. doi: 10.1177/0163278702025001006

Stewart, L. A., Tierney, J. F., & Clarke, M. (2008). Reviews of individual patient data. In P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (pp. 547-558). Chichester, England: Wiley-Blackwell.

Taylor, P., Hutton, P., & Dudley, R. (2014). Rationale and protocol for a systematic review and meta-analysis on reduced data gathering in people with delusions. *Systematic Reviews, 3*(1), 44. doi: 10.1186/2046-4053-3-44

Team, R. C. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Terrin, N., Schmid, C. H., & Lau, J. (2005). In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *Journal of Clinical Epidemiology, 58*(9), 894-901. doi: 10.1016/j.jclinepi.2005.01.006

Thompson, S. G., & Higgins, J. P. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine, 21*(11), 1559-1573. doi: 10.1002/sim.1187

van der Leer, L., Hartig, B., Goldmanis, M., & McKay, R. (2015). Delusion-proneness and 'jumping to conclusions': relative and absolute effects. *Psychological Medicine, 45*(6), 1253-1262. doi: 10.1017/s0033291714002359

van der Leer, L., & McKay, R. (2014). “Jumping to conclusions” in delusion-prone participants: an experimental economics approach. *Cognitive Neuropsychiatry*, *19*(3), 257-267. doi: 10.1080/13546805.2013.861350

Van Os, J., Hanssen, M., Bijl, R. V., & Ravelli, A. (2000). Strauss (1969) revisited: a psychosis continuum in the general population? *Schizophrenia Research, 45*, 11-20. doi: 10.1016/s0920-9964(00)90323-2

van Os, J., Linscott, R. J., Myin-Germeys, I., Delespaul, P., & Krabbendam, L. (2009). A systematic review and meta-analysis of the psychosis continuum: Evidence for a psychosis proneness–persistence–impairment model of psychotic disorder. *Psychological Medicine, 39*(2), 179-195. doi: 10.1017/S0033291708003814

Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48.

Wallace, B. C., Dahabreh, I. J., Trikalinos, T. A., Lau, J., Trow, P., & Schmid, C. H. (2012). Closing the gap between methodologists and end-users: R as a computational back-end. *Journal of Statistical Software, 49*(5), 1-15.

Warman, D. M. (2008). Reasoning and delusion proneness: Confidence in decisions. *Journal of Nervous and Mental Disease, 196*(1), 9-15. doi: 10.1097/NMD.0b013e3181601141

Warman, D. M., Lysaker, P. H., Martin, J. M., Davis, L., & Haudenschield, S. L. (2007). Jumping to conclusions and the continuum of delusional beliefs. *Behaviour Research and Therapy, 45*(6), 1255-1269. doi: 10.1016/j.brat.2006.09.002

Warman, D. M., & Martin, J. M. (2006). Jumping to conclusions and delusion proneness: the impact of emotionally salient stimuli. *Journal of Nervous and Mental Disease, 194*(10), 760-765. doi: 10.1097/01.nmd.0000239907.83668.aa

White, L. O., & Mansell, W. (2009). Failing to ponder? Delusion-prone individuals rush to conclusions. *Clinical Psychology & Psychotherapy, 16*(2), 111-124. doi: 10.1002/cpp.607

Whitman, J. C., Menon, M., Kuo, S. S., & Woodward, T. S. (2013). Bias in favour of self-selected hypotheses is associated with delusion severity in schizophrenia. *Cognitive Neuropsychiatry, 18*(5), 376-389. doi: 10.1080/13546805.2012.715084

Woodward, T. S., Moritz, S., Cuttler, C., & Whitman, J. C. (2006). The contribution of a cognitive bias against disconfirmatory evidence (BADE) to delusions in schizophrenia. *Journal of Clinical and Experimental Neuropsychology, 28*(4), 605-617. doi: 10.1080/13803390590949511

Woodward, T. S., Munz, M., LeClerc, C., & Lecomte, T. (2009). Change in delusions is associated with change in "jumping to conclusions". *Psychiatry Research, 170*(2-3), 124-127. doi: 10.1016/j.psychres.2008.10.020

Ziegler, M., Rief, W., Werner, S.-M., Mehl, S., & Lincoln, T. M. (2008). Hasty decision-making in a variety of tasks: does it contribute to the development of delusions? *Journal of Psychology & Psychotherapy, 81*(3), 237-245. doi: 10.1348/147608308X297104

Chapter Three: Analytic cognitive style, not delusional ideation, predicts data gathering in a large beads task study

3.1 Abstract

Cognitive models of delusions propose that delusional and delusion-prone individuals gather unusually little evidence before forming beliefs. Crucial evidence for this "jumping to conclusions" (JTC) bias has been provided by studies using the "beads task" data gathering paradigm. However, a cognitive model of normal reasoning in this task has yet to be developed. Consequently, the cognitive dysfunctions responsible for the JTC bias are poorly understood. Recently, it has been suggested that cognitive dysfunctions associated with delusions could be specified with increased precision in terms of dual-process theories of reasoning. Inspired by this suggestion, we examine the cognitive mechanisms subserving data gathering in the beads task. In the largest beads task study to date (n = 558), we find that increased "analytic cognitive style" predicts greater data gathering, even after controlling for potential confounds, but delusional ideation does not.

3.2 Introduction

Research suggests that reasoning biases play an important role in the formation and persistence of delusional beliefs (Connors & Halligan, 2015; Garety & Freeman, 1999, 2013). In a pioneering study, Huq, Garety, and Hemsley (1988) adapted the "beads task" data-gathering paradigm (Phillips, Hayes, & Edwards, 1966) to examine the relationship between delusions and data gathering. They presented participants with two jars filled with coloured beads of complementary ratios: a mostly pink jar (85 pink beads; 15 green beads) and a mostly green jar (85 green beads; 15 pink beads). After these jars had been hidden, participants were shown a sequence of beads that were ostensibly being drawn from one of these jars (the sequence was actually prespecified). After each bead was revealed participants were asked whether they wanted to see another bead or make a decision about which jar the beads were being drawn from. Participants with delusions requested significantly fewer beads than psychiatric and healthy controls. Dozens of beads task studies have subsequently been published, many of which have replicated this result (Garety & Freeman, 1999, 2013). Consequently, it has been argued that delusional and delusion prone individuals have a "jumping to conclusions" (JTC) bias that results in beliefs being formed on the basis of unusually little evidence. This JTC bias has featured prominently in contemporary cognitive models of delusions and meta-cognitive training programmes designed to target delusional reasoning (Garety et al., 2011; Garety, Kuipers, Fowler, Freeman, & Bebbington, 2001; Moritz et al., 2014).

A fundamental goal of cognitive neuropsychiatry is to develop cognitive models that explain psychiatric symptoms in terms of dysfunctions in specific information-processing procedures (Coltheart, 2007). Unfortunately, little progress has been made in identifying the specific dysfunctions responsible for the association between the JTC bias and delusions. One difficulty is that a cognitive model of normal reasoning in the beads task has not yet been developed. This lack of specificity about dysfunctions curtails the explanatory power of the

JTC bias. In fact, some scholars have argued that the JTC bias has no explanatory power whatsoever: "portrayal of people with delusions as having a JTC bias is a redescription rather than an explanation" (Corlett & Fletcher, 2014, p. 399). Nevertheless, we suggest that there is scope for using the beads task to identify specific cognitive dysfunctions. In particular, a productive first step would be to characterise typical variation in data gathering in the general population in terms of a theory of normal reasoning.

A substantial body of evidence supports the existence of two relatively distinct reasoning processes (Evans, 2010; Evans & Stanovich, 2013; Kahneman, 2003, 2011; Stanovich, 2011). One of these, Type 1 or "intuitive" processing, does not require working memory, and is typically fast, high capacity, parallel, unconscious, automatic, associative, and independent of cognitive ability. The other, Type 2 or "analytic" processing, requires working memory, and is typically slow, low capacity, serial, conscious, voluntary, deliberative, and dependent on cognitive ability. Type 1 processing provides fast intuitive responses that guide behaviour unless slower Type 2 monitoring processes intervene and offer an alternative response (Evans, 2007).

Dual-process theories of reasoning have been investigated using a variety of reasoning tasks. Consider, for example, the widely studied "bat and ball problem" included in the Cognitive Reflection Test (Frederick, 2005). "A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?" Most people respond to this problem with the first answer that comes to mind: 10 cents. However, thinking more carefully reveals that this intuitively appealing answer is incorrect. If the ball cost 10 cents then the bat must cost $1.10, which means that the total cost is $1.20, not $1.10. Performance on this task and others like it, provides evidence for substantial variation in "analytic cognitive style" that is related to, but dissociable from, cognitive ability, and is defined as a propensity to critically evaluate outputs from Type 1 intuitive processing and engage in effortful Type 2 analytic processing (Stanovich, 2011; Stanovich & West, 2008; Toplak, West, & Stanovich, 2011, 2013).

There exist striking parallels between dual-process theories of normal reasoning and cognitive models of delusional reasoning. According to dual-process theories, Type 1 processing acts as a "machine for jumping to conclusions" (Kahneman, 2011, p. 79) that is "radically insensitive to both the quality and the quantity of the information that gives rise to impressions and intuitions" (p. 86). Similarly, cognitive models of delusional reasoning posit that a JTC bias plays an important role in delusion formation because "anomalous or ambiguous information is rapidly appraised and a delusion is formed on the basis of limited evidence" (Garety & Freeman, 2013, p. 327). Drawing on such parallels, a number of scholars have recently suggested that the cognitive style of people with delusions might be characterised by an over-reliance on Type 1 processing (Aimola Davies & Davies, 2009; Freeman, Evans, & Lister, 2012; Freeman & Garety, 2014; Freeman, Lister, & Evans, 2014; Garety et al., 2015; Speechley, Murray, McKay, Munz, & Ngan, 2010; Speechley & Ngan, 2008; Speechley, Whitman, & Woodward, 2010; Speechley, Woodward, & Ngan, 2013)[9]. In other words, the suggestion is that delusional reasoning is negatively associated with analytic cognitive style.

Although the relationship between data gathering and delusions has been examined in dozens of beads task studies (Garety & Freeman, 1999, 2013), surprisingly little research has examined whether data gathering in the beads task can be predicted using dual-process theories of reasoning. Nevertheless, a recent study by Brosnan, Hollinworth, Antoniadou, and Leton (2014) is relevant. They found that performance in the Cognitive Reflection Test is positively correlated with data gathering in the beads task, which provides support for a connection between data gathering and Type 2 processing. Nevertheless, this study leaves

[9] Some of these authors refer to "System 1" and "System 2", or "Stream 1" and "Stream 2". However, "Type 1" and "Type 2" are now recommended as the most appropriate technical terms (see Evans & Stanovich, 2013, p. 224-226) and are used throughout this paper.

two important issues unaddressed. First, being a study designed to test hypotheses about empathising and systematising, not delusional reasoning, a measure of delusional ideation was not included. Consequently, the possibility that delusional ideation and performance in the Cognitive Reflection Test might independently predict data gathering could not be explored. Second, this study did not control for raw cognitive ability, which is associated with performance in many reasoning tasks, including the Cognitive Reflection Test (Stanovich, 2011; Stanovich & West, 2008; Toplak et al., 2011, 2013). Consequently, this study could not determine whether analytic cognitive style, or cognitive ability, or both, was responsible for the association.

In the present study we investigated whether analytic cognitive style is associated with data gathering in the beads task. Our primary hypothesis was that increased analytic cognitive style predicts greater data gathering. In addition, we aimed to replicate the finding that increased delusional ideation predicts less data gathering (Ross, McKay, Coltheart, & Langdon, 2015). Since we are interested in providing evidence for causal theories using correlational data, we controlled for a variety of potential confounds. In particular, we controlled for cognitive ability, traditional religiosity, and paranormal belief because they have been associated with cognitive style, or data gathering, or both (Gervais & Norenzayan, 2012; Irwin, Drinkwater, & Dagnall, 2014; Lawrence & Peters, 2004; Pennycook, Cheyne, Barr, Koehler, & Fugelsang, 2013, 2014; Pennycook, Cheyne, Koehler, & Fugelsang, 2013; Pennycook, Cheyne, Seli, Koehler, & Fugelsang, 2012; Peters, Day, McKenna, & Orbach, 1999; Shenhav, Rand, & Greene, 2012).

3.2 Methods

3.2.1 Participants

Participants were recruited through *Mechanical Turk*, an online marketplace where "workers" can sign up for paid tasks, including psychological studies (Buhrmester, Kwang, & Gosling, 2011). Only people with a strong track record of completing tasks satisfactorily (a "HIT approval rating" of greater or equal to 95%) and a USA-based *Mechanical Turk* user account were eligible to respond to the advertisement. Participation was voluntary and participants received $US2.20 as remuneration. Sessions lasted approximately 45 minutes.

We included two instructional manipulation checks to ensure that participants were paying attention to tasks. Following the advise of Oppenheimer, Meyvis, and Davidenko (2009) we did not exclude participants who failed these checks on their first attempt. Rather, these participants were informed that they had made an error and were asked to attempt the instructional manipulation check again. If a participant passed on their second attempt, we took this as evidence that their attention had been refocused, and their data were retained for further analysis. If a participant failed on their second attempt, we took this as evidence that they were not motivated to follow simple instructions, and their data were excluded from the analysis.

Surveys that were incomplete were excluded from analysis. Complete surveys were subjected to the following sequential screening criteria (the number of participants failing a given criterion are shown in parentheses): they had a USA IP address (22); they had an IP address that did not match that of any other participant (24); they indicated that they are a fluent speaker of English (5); they indicated that they had not answered questions randomly and had given answers that reflect their true beliefs (13); they indicated that they had not consulted the Internet or other people to get answers to reasoning problems (20); they passed both Instructional Manipulation checks on their first or second attempt (46); they indicated that they are 18 or older (1; the minimum age for being allowed to open a *Mechanical Turk* account is 18); they asked to see fewer than 50 beads during the beads task (12; the maximum number possible in our study is 50—see justification below) After screening, 558 participants were retained for analysis.

3.3 Materials

3.3.1 Beads task

Participants were shown an image of two jars of beads and task instructions based on those of Garety *et al.* (2005). One jar was labelled as being the "mainly red jar" and was pictured as having 60 red beads and 40 blue beads, and the other jar was labelled as being the "mainly blue jar" and was pictured as having 40 red beads and 60 blue beads. We used the "hard version" of the beads task (complementary ratios of 60:40 and 40:60) rather than the "easy version" (complementary ratios of 85:15 and 15:85) because we wanted high posterior probabilities for the correct jar to appear relatively late in the sequence (see Maher [2004] for a discussion of difficulties associated with interpreting results when the easy version of the beads task is used). After the images of the jars had been removed, participants were shown a sequence of beads apparently being drawn from one of the jars with replacement. In reality, the sequence of beads was prespecified and identical for all participants (b = blue; r = red): b r r b b r b b b r b b b b r r b r r b b r b b r b b b r r b r r b b b b r b b r r r r b b r b b b. The first 20 beads followed a widely used sequence from Garety *et al.* (2005) that stopped at 20 beads. We added an additional 30 beads to the sequence (maintaining roughly the same ratio of blue to red beads throughout the sequence) to provide participants with the opportunity to see a substantial number of beads. After each draw, participants were asked if they would like to decide which jar the beads were being drawn from or if they would like to see another bead. As per standard procedure, data gathering was operationalised as the number of beads a participant asked to see before making a decision—i.e. the "draws to decision". Requesting a vast number of beads suggests that a participant either did not pay close attention to the sequence of beads or interpreted the task instructions differently to other participants. Consequently, we removed participants who asked to see all 50 beads prior to analysis. In total, only 12 participants (i.e. 2.1% of otherwise eligible participants) were removed on this basis.

3.3.2 Analytic cognitive style

We used four tests to measure analytic cognitive style. First, the 3-item Cognitive Reflection Test (CRT3; Frederick, 2005). The CRT3 consists of simple mathematical problems, including the bat and ball problem, that generate implicit misleading conclusions. Correct responses were summed to create a CRT3 score. The CRT3 has been used in many studies and there is evidence that some participants in *Mechanical Turk* studies have been repeatedly exposed to it (Chandler, Mueller, & Paolacci, 2013). To explore the possibility that prior exposure to the CRT3 might bias results, participants were asked if they had seen any of the items previously and, if so, whether they had been provided with the correct answer.

Second, the 4-item Cognitive Reflection Test (CRT4; Toplak et al., 2013). Like the CRT3, the CRT4 consists of simple mathematical problems that generate implicit misleading conclusions. However, the CRT4 has two advantages. First, the CRT4 is not currently widely known, so responses are very unlikely to be biased by previous exposure. Second, the CRT3 is very difficult, with students at elite universities frequently providing incorrect responses (Frederick, 2005), which suggests that a floor effect might be evident in many populations. The CRT4 is considerably easier (Toplak et al., 2013). Correct responses were summed to create a CRT4 score.

Third, a 15-item "heuristics and biases" battery (Toplak et al., 2011). This battery was designed to reflect important aspects of rational thought. For example, one questions asks, "When playing slot machines, people win approximately 1 in every 10 times. Julie, however, has just won on her first three plays. What are her chances (out of 10) of winning the next time she plays?" Correct responses were summed to create a heuristics and biases score.

Fourth, an 8-item syllogistic reasoning test (De Neys & Franssens, 2009). Four items are "conflict problems" that have conclusions in which deductive logic was in conflict with believability (two items had unbelievable-valid conclusion, and two items had believable-invalid concussions); and four items are "non-conflict problems" that have conclusions in which deductive logic was consistent with believability (two problems had unbelievable-invalid conclusions, and two problems had believable-valid conclusions). For example, consider one of the conflict problems, “Premise 1: All vehicles have wheels. Premise 2: A boat is a vehicle. Conclusion: Therefore, a boat has wheels. Assume that the two premises are true. Does the conclusion follow logically from the two premises?” The problem is logically valid but has an unbelievable conclusion (boats do not have wheels). Correct responses to the four conflict problems were summed to create a conflict syllogism score.

An overall analytic cognitive style score was calculated by summing scores from the CRT3, the CRT4, the heuristics and biases battery, and the conflict syllogisms together (minimum possible score 0; maximum possible score 26). We found that the scale had acceptable internal consistency: Cronbach’s $\alpha = 0.81$.

3.3.3 Cognitive ability

Solving any reasoning task requires both cognitive ability and analytic cognitive style. Nevertheless, different tasks pose different cognitive challenges (Stanovich, 2011; Toplak et al., 2011, 2013). To solve the bat and ball problem, for example, requires a high level of analytic processing to inhibit and override the intuitive incorrect response, and some degree of numeracy (i.e., cognitive ability) for the simple arithmetic. By contrast, basic numeracy problems do not present an incorrect intuitive lure, so little analytic cognitive style is needed to answer correctly. Consequently, the bat and ball problem is more strongly reflective of analytic cognitive style than basic numeracy problems. Following earlier research, we use tasks that predominantly measure cognitive ability as controls and examine how much additional variation in draws to decision is explained by performance in tasks that require some degree of analytic cognitive style (Barr, Pennycook, Stolz, & Fugelsang, 2015a; Cheyne & Pennycook, 2013; Pennycook, Cheyne, Barr, et al., 2013; Pennycook et al., 2014;

Pennycook, Cheyne, Koehler, et al., 2013; Pennycook et al., 2012; Toplak et al., 2011, 2013; Trippas, Pennycook, Verde, & Handley, in press).

We used three tasks to measure cognitive ability. First, correct responses to the four non-conflict problems from the syllogistic reasoning test discussed above were summed to create a non-conflict syllogism score.

Second, the Wordsum test (Huang & Hauser, 1998). This verbal intelligence test comprises of 10 multiple choice vocabulary questions in which participants are asked to identify which of five words comes closest in meaning to a target word. For example, one of the target words is *animosity* and the five options are *hatred, animation, disobedience, diversity*, and *friendship*. The Wordsum test correlates well with full-scale measure of intelligence (Huang & Hauser, 1998), and has been used in 16 General Social Surveys (Davies & Smith, 1994) and numerous psychological, sociological, and political science studies (Malhotra, Krosnick, & Haertel, 2007).

Third, a 3-item basic numeracy test (Schwartz, Woloshin, Black, & Welch, 1997). For example, one of the questions is “Imagine that we flip a fair coin 1,000 times. What is your best guess about how many times the coin would come up heads in 1,000 flips?” Scores on this test are strongly associated with scores on longer 7-item numeracy test (Lipkus, Samsa, & Rimer, 2001). We opted for the shorter version to minimise the length of the study.

An overall cognitive ability score was calculated by summing scores from the Wordsum test and the numeracy test together (minimum possible score 0; maximum possible score 17). We found that the scale had acceptable internal consistency: Cronbach’s $\alpha = 0.63$.

3.3.4 Delusional ideation

The 21-item Peters *et al.* Delusions Inventory (PDI; Peters, Joseph, Day, & Garety, 2004; Peters, Joseph, & Garety, 1999) was used to measure delusional ideation. Participants are asked if they have ever had any of 21 delusion-like experiences. For example, one item asks, “Do you ever feel as if things in magazines or on TV were written especially for you?” For each item endorsed participants are asked to rate the associated distress, preoccupation, and conviction. The PDI has been used in numerous studies, including at least 22 beads task studies (Ross et al., 2015). We found that the scale had acceptable internal consistency: Cronbach’s $\alpha = 0.75$.

3.3.5 Religious belief and participation

A 9-item religious belief scale (Pennycook et al., 2014; Pennycook et al., 2012) was used to measure conventional religious beliefs. And a 5-item religious participation scale (Pennycook et al., 2014; Pennycook et al., 2012) was used to assess the frequency of participation in conventional religious activities. We found that the scales had good internal consistency: Cronbach’s $\alpha = 0.95$ and $\alpha = 0.89$ for religious belief and religious engagement respectively.

3.3.6 Paranormal ideation

The 25-item Revised Paranormal Belief Scale (Tobacyk, 2004; Tobacyk & Milford, 1983) was used to measure paranormal belief. Three religious items were removed because they were made redundant by the religious belief scale. We found that the scale had good internal consistency: Cronbach’s $\alpha = 0.95$.

3.3.7 Demographic variables

Participants were asked to report their gender (Male 63.9%, Female 36.1%); age in years (Mean = 30.2); highest level of education [1 = None (0%), 2 = some high school (1.3%), 3 = high school (9.8%), 4 = technical trade or vocational training (5.7%), 5 = some college, no degree (41.7%), 6 = Bachelor’s degree (29.2%), 7 = Master’s degree (10.0%), 8 = Professional degree (0.9%), 9 = Doctoral degree (1.4%)]; and an estimate of total family income level from all sources before taxes for 2013 [1 = less than $10,000 (7.7%), 2 =

$10,000 to under $20,000 (13.1%), 3 = $20,000 to under $30,000 (14.7%), 4 = $30,000 to under $40,000 (11.8%), 5 = $40,000 to under $50,000 (11.1%), 6 = $50,000 to under $75,000 (15%), 7 = $75,000 to under $100,000 (11.8%), 8 = $100,000 to under $150,000 (10.6%), 9 = $150,000 to under $250,000 (2.7%), 10 = $250,000 or more (1.6%)].

3.3.8 Procedure

The study was conducted using *Qualtrics* (Provo, UT, USA: Qualtrics). The order of presentation of tasks was the same for all participants and was as follows:

1. A declaration of informed consent.
2. The first manipulation check.
3. The beads task.
4. The PDI.
5. The CRT4.
6. The CRT3.
7. The second manipulation check.
8. The heuristics and biases battery.
9. The syllogistic reasoning test.
10. The Wordsum test.
11. The numeracy test.
12. The religious belief and religious participation scales.
13. The paranormal belief scale.
14. Pilot questionnaires that are not relevant to the present study.
15. Demographic questions.
16. Questions about honesty and accuracy of responses.
17. Debrief.

To facilitate interpretation of results and comparison with other studies, we converted scores on tests and psychometric measures to Percentage of Maximum Possible (POMP) scores (Cohen, Cohen, Aiken, & West, 1999), where POMP score = [(*observed score – minimum possible*)/(*maximum possible – minimum possible*)] × 100. Thus, regardless of the measure, the score on these scales vary from 0 to 100. The single exception was the beads task, which has an arbitrary upper limit and is best interpreted using raw scores using the draws to decision measure.

3.4 Results

A frequency distribution of draws to decision is shown in Figure 3.1. This distribution shows that the vast majority of participants asked to see 20 or fewer beads – the typical upper limit for beads task studies.

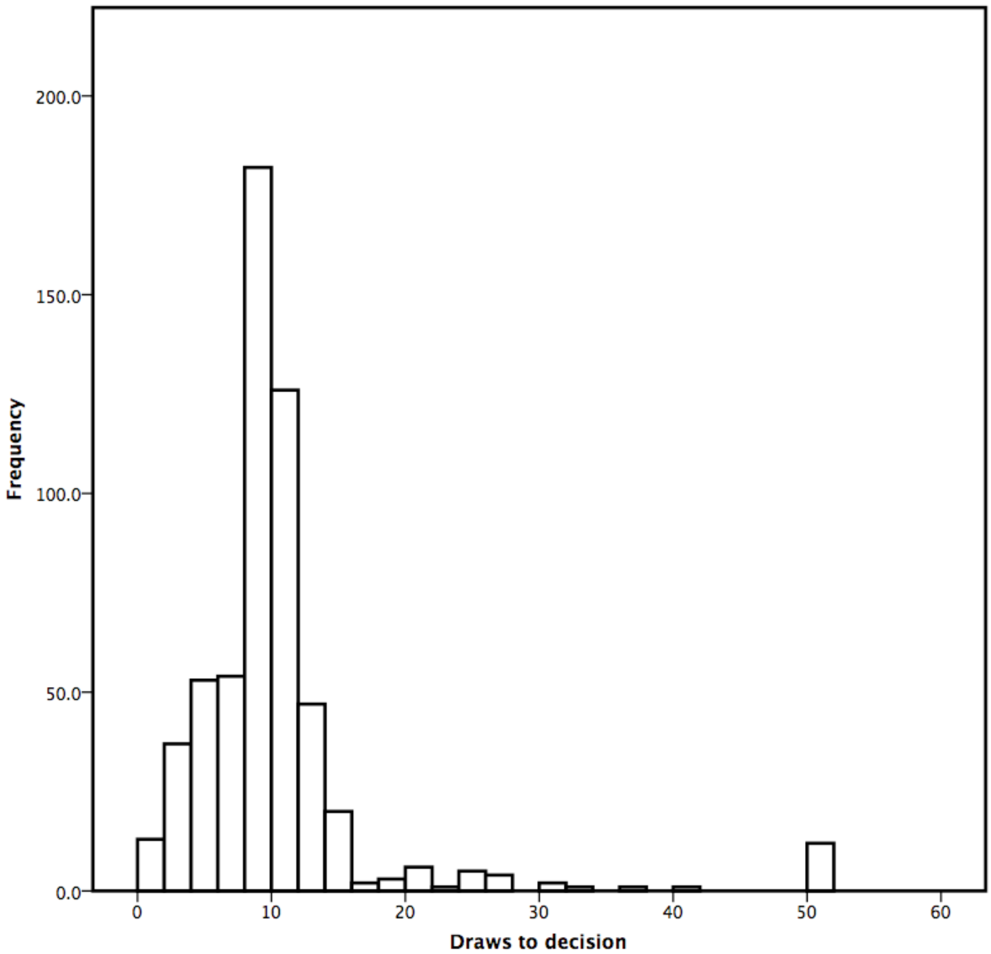


Figure 3.1 Frequency distribution of draws to decision.

Descriptive statistics are reported in Table 3.1. A mixture of categorical, ordered, and continuous variables were analysed. Furthermore, some of the ordered and continuous variables were not normally distributed. For these reasons we used Spearman's rank-order correlations (r_s). All five analytic cognitive style measures (CRT4, CRT3, heuristics and biases battery, and conflict syllogisms) are positively associated with each other (all *p*-values < .01); and all three cognitive ability measures (non-conflict syllogisms, Wordsum test, and numeracy test) are positively associated with each other (all *p*-values < .01). These findings suggest that it is appropriate to sum scores from individual tests to form composite measures of analytic cognitive style and cognitive ability, as has been done in earlier studies (Barr et al., 2015a; Barr, Pennycook, Stolz, & Fugelsang, 2015b; Cheyne & Pennycook, 2013; Pennycook, Cheyne, Barr, et al., 2013; Pennycook et al., 2014; Pennycook, Cheyne, Koehler, et al., 2013; Pennycook et al., 2012; Trippas et al., in press).

Spearman's rank-order correlations (r_s) are reported in Table 3.2. With respect to the hypotheses being examined, the most important correlations are between draws to decision, analytic cognitive style, and PDI. Draws to decision is positively correlated with analytic cognitive style, but not PDI. In addition, it is noteworthy that draws to decision is positively correlated with cognitive ability, and education; and is negatively correlated with paranormal belief. The extent to which draws to decision can be predicted by analytic cognitive style, PDI, and the other variables independently of each other is explored using hierarchical multiple regression.

Hierarchical multiple regression analysis is reported in Table 3.3. Variables were entered into the model in the follow order: first step – demographic variables (age, gender, education and income); second step – PDI; third step – paranormal belief; fourth step – religious belief and religious engagement; fifth step – cognitive ability; and sixth step – analytic cognitive style. Visual inspection of a plot of predicted values of draws to decision against residuals indicated that the linear model's assumption of normality of the error

Table 3.1 Descriptive statistics.

	Mean	Minimum	Maximum	SD	Skew	Kurtosis
Draws to decision	9.17	1	41	4.65	2.18	9.79
PDI	17.30	0	60	10.87	0.95	1.10
CRT4	42.20	0	100	31.38	0.35	-0.85
CRT3	35.07	0	100	35.77	0.51	-1.10
Heurstics and biases	49.43	7	93	17.09	0.21	-0.27
Conflict syllogisms	61.96	0	100	38.01	-0.41	-1.32
Analytic cognitive style	48.59	8	96	19.30	0.20	-0.76
WordSum	76.29	0	100	16.34	-0.83	1.79
Numeracy Test	68.46	0	100	29.16	-0.51	-0.66
Cognitive Ability	74.48	0	100	15.84	-0.66	0.82
Religious Belief	53.50	0	100	31.40	-0.24	-1.12
Religious Participation	27.50	0	100	27.03	1.04	0.14
Paranormal belief	30.55	0	100	21.25	0.34	-0.69
Age	30.21	18	76	9.86	1.16	1.20
Sex	0.36	0	1	0.48	0.58	-1.67
Education	53.63	13	100	15.16	-0.10	0.86
Income	41.96	0	100	26.01	0.13	-0.99

Note. Sex (Male = 0; Female = 1); *SE* of Skew = 0.1; *SE* of Kurtosis = 0.21; $n = 558$.

Table 3.2 Spearman's rank-order correlations.

Variable	PDI	CA	ACS	RB	RP	PB	Age	Sex	Edu	Income
DTD	-.016	.140**	.259**	-.046	-.056	-.105*	-.078	.037	.143**	.034
PDI		-.092*	-.073	.298**	.260**	.436**	-.155**	-.023	-.141**	-.133**
CA			.541**	-.189**	-.085*	-.182**	.088*	.090*	.239**	.123**
ACS				-.224**	-.114**	-.276**	.027	.334**	.287**	.139**
RB					.747**	.367**	.032	-.187**	-0.05	-.033
RP						.124**	.054	-.089*	-.014	-.011
PB							.056	-.239**	-.161**	-.226**
Age								-0.04	.241**	.058
Sex									-.008	.094*
Edu										.211**

Note. DTD = draws to decision; PDI = Peters *et al.* Delusions Inventory; CA = Cognitive ability; ACS = Analytic cognitive style; RB = Religious belief; RP = Religious participation; PB = Paranormal belief; Sex (Male = 0; Female = 1); Edu = Education. $*p \leq .05$ and $**p \leq .01$, two-tailed tests; $n = 558$.

distribution was violated. For this reason we re-ran the analysis using bootstrap resampling implemented in *SPSS* version 21 (Armonkm, NY, USA: IBM Corp.) to calculate beta coefficients, beta standard errors, bias corrected confidence intervals, and p-values for each of the predictors that are robust to this departure from this assumption of the linear model (Field, 2013).

At step 1, age and education is a significant predictor of draws to decision. At step 2, age and education are significant predictors, but the newly introduced PDI is not. At step 3, education and the newly introduced paranormal belief are significant predictors. At step 4, education is a significant predictor, but the newly introduced religious belief and religious participation are not. At step 5, age, education, and the newly introduced cognitive ability are significant predictors. At step 6, the newly introduced analytic cognitive style is the only significant predictor. In summary, the hierarchical regression analysis indicates that analytic cognitive style makes a contribution to predicting draws to decision in the beads task. Notably, analytic cognitive style dominates prediction to such an extent that the three variables that are significant predictors in the previous step (age, education, and cognitive ability) are no longer significant once analytic cognitive style has been introduced into the regression, which suggests that other variables predicted draws to decision only to the extent that they are associated with analytic cognitive style.

Table 3.3 Hierarchical multiple regression analyses predicting draws to decision.

							95% C.I. for B	
		B	Std. Error	Beta	t	Sig.	Lower	Upper
1	(Constant)	7.69	0.79		8.606	0.00	6.07	9.24
	Age	-0.04	0.02	-0.092	-2.137	0.05	-0.08	0.00
	Sex	0.18	0.41	0.019	0.446	0.65	-0.58	0.96
	Education	0.05	0.01	0.165	3.79	0.00	0.02	0.08
	Income	0.00	0.01	0	0.011	0.99	-0.02	0.02
2	(Constant)	8.11	0.93		7.967	0.00	6.20	9.98
	Age	-0.05	0.02	-0.096	-2.222	0.04	-0.09	0.00
	Sex	0.17	0.41	0.018	0.416	0.67	-0.59	0.95
	Education	0.05	0.01	0.161	3.7	0.00	0.02	0.08
	Income	0.00	0.01	-0.003	-0.068	0.96	-0.02	0.02
	PDI	-0.02	0.02	-0.037	-0.854	0.37	-0.05	0.02
3	(Constant)	8.63	0.98		8.234	0.00	6.81	10.48
	Age	-0.04	0.02	-0.081	-1.856	0.08	-0.08	0.01
	Sex	-0.02	0.44	-0.002	-0.052	0.96	-0.81	0.76
	Education	0.05	0.01	0.149	3.398	0.00	0.02	0.07
	Income	0.00	0.01	-0.017	-0.399	0.68	-0.02	0.01
	PDI	0.00	0.02	0.005	0.104	0.91	-0.04	0.04
	Paranormal Belief	-0.02	0.01	-0.099	-1.997	0.05	-0.05	0.00
4	(Constant)	8.70	0.98		8.14	0.00	6.91	10.49
	Age	-0.04	0.02	-0.082	-1.861	0.07	-0.08	0.00
	Sex	-0.04	0.43	-0.004	-0.089	0.93	-0.82	0.75
	Education	0.05	0.01	0.15	3.399	0.00	0.02	0.07
	Income	0.00	0.01	-0.016	-0.369	0.69	-0.02	0.01
	PDI	0.00	0.02	0.003	0.066	0.95	-0.04	0.04
	Paranormal Belief	-0.02	0.01	-0.089	-1.601	0.11	-0.05	0.00
	Religious Belief	0.00	0.01	-0.028	-0.415	0.61	-0.03	0.02
	Religious Participation	0.00	0.01	0.021	0.333	0.73	-0.02	0.03
5	(Constant)	7.00	1.27		5.243	0.00	4.66	9.34
	Age	-0.04	0.02	-0.089	-2.025	0.05	-0.08	0.00
	Sex	-0.07	0.43	-0.007	-0.159	0.88	-0.85	0.75
	Education	0.04	0.01	0.132	2.958	0.00	0.02	0.07
	Income	0.00	0.01	-0.023	-0.537	0.59	-0.02	0.01
	PDI	0.00	0.02	-0.002	-0.043	0.96	-0.04	0.04
	Paranormal Belief	-0.02	0.01	-0.078	-1.404	0.14	-0.04	0.01
	Religious Belief	0.00	0.01	-0.019	-0.291	0.71	-0.02	0.02
	Religious Participation	0.01	0.01	0.026	0.419	0.66	-0.02	0.03
	Cognitive Ability	0.03	0.01	0.093	2.105	0.03	0.00	0.05
6	(Constant)	7.16	1.27		5.474	0.00	4.74	9.56
	Age	-0.04	0.02	-0.08	-1.844	0.07	-0.08	0.01
	Sex	-0.78	0.45	-0.08	-1.773	0.09	-1.57	0.02
	Education	0.03	0.01	0.08	1.79	0.06	0.00	0.05
	Income	0.00	0.01	-0.022	-0.525	0.58	-0.02	0.01
	PDI	-0.01	0.02	-0.013	-0.274	0.75	-0.05	0.03
	Paranormal Belief	-0.01	0.01	-0.052	-0.958	0.33	-0.04	0.01
	Religious Belief	0.00	0.01	-0.005	-0.077	0.93	-0.02	0.02
	Religious Participation	0.00	0.01	0.026	0.423	0.67	-0.02	0.03
	Cognitive Ability	-0.01	0.02	-0.021	-0.426	0.67	-0.04	0.02
	Analytic Cognitive Style	0.06	0.01	0.264	4.967	0.00	0.04	0.09

Note. Sex (Male = 0; Female = 1); PDI = Peters *et al.* Delusions Inventory; paranormal belief; religious belief and religious engagement; cognitive ability. 95% bias corrected and accelerated confidence intervals and standard errors are based on 1000 bootstrap samples.

R^2 = .03 for Step 1 (p < .01); ΔR^2 = .00 (p = .39) for Step 2; ΔR^2 = .01 (p = .05) for Step 3; ΔR^2 = .00 (p = .92) for Step 4; ΔR^2 = .01 (p = .04) for Step 5; ΔR^2 = .04 (p < .01) for Step 6; n = 558.

Quite a few participants indicated that they had previously seen items from the CRT3. The mean number of items was 1.45, with 31.4% of participants reporting having seen at least one item. Furthermore, 13.8% of participants reported having been provided with the correct answer to at least one item. Because previous exposure to the CRT3 might have had an influence on results, we investigated the robustness of the results by re-running step 6 of the regression with the CRT3 removed from the measure of Analytic Cognitive Style. We found that the results were very similar. Consequently, we focus on the analysis that included the CRT3 in the discussion.

3.5 Discussion

The present study provides evidence for a relationship between analytic cognitive style, defined as a propensity to critically evaluate the output of intuitive processing and engage in effortful analytic reasoning, and data gathering in the beads task. Draws to decision measure was positively correlated with analytic cognitive style and this association was stronger than that between draws to decision and any other variable. Nonetheless, the tasks that were used to measure analytic cognitive style also depend on cognitive ability (Stanovich, 2011; Stanovich & West, 2008; Toplak et al., 2011, 2013), so it is possible that cognitive ability, rather than analytic cognitive style, is associated with data gathering in the beads task. The regression analysis provided evidence against this interpretation because performance in tasks that measure analytic cognitive style predicted variation in draws to decision after controlling for performance in tasks that measure cognitive ability alone.

The present study did not find evidence for a relationship between delusional ideation and data gathering. There was no statistically significant correlation between PDI score and draws to decision. Likewise, regression analysis provided no evidence that PDI score predicts draws to decision. These results differ from that of a recent meta-analysis of beads task studies that found a negative correlation between delusional ideation and data gathering (Ross et al., 2015). Nevertheless, we suggest that the null result from the present study can be harmonised with the result of this meta-analysis. Although the meta-analysis provided evidence for a statistically significant association, the effect size was very small (r_s = -0.10, 95% CI [-0.17, -0.03]), which suggests that an association may only have been detected because the sample size of the meta-analysis was very large (n = 2,237). Consequently, if the results of the meta-analysis capture the true effect size, a sample of 558 participants might be insufficient to reliably detect the association. Alternatively, it is possible that the present study did not find an association between PDI and DTD due to limitations in the study design itself. One possibility is that DTD is a suboptimal measure of data gathering because the difference between, say, 35 and 40 beads is given the same weight as the difference between 5 and 10 beads. Future research could probe subjective prior and posterior probabilities to provide a stronger measure of data gathering. In any respect, if there exists a genuine association between data gathering and delusional ideation, the present results suggest that delusional ideation explains considerably less variation in data gathering than analytic cognitive style.

Although analytic cognitive style predicts data gathering in the beads task, it is not clear why. Superficially, the beads task resembles typical tasks from the reasoning literature. However, there is a crucial difference. Reasoning tasks such as the bat and ball problem evoke an intuitively appealing, but incorrect, response that can be substituted with a normatively correct response by careful thinking. By contrast, although the term “jumping to conclusions” suggests that data gathering has been *insufficient* and a decision has been reached *prematurely*, the JTC bias studied using the beads task is a relative effect. In standard versions of the beads task there is, in fact, no normatively correct point at which a rational individual *should* make a decision (van der Leer, Hartig, Goldmanis, & McKay, 2015; van der Leer & McKay, 2014). Consequently, making inferences about the cause of the association between analytic cognitive style and draws to decision is a challenge. Nevertheless, we suggest that there are at least two plausible hypotheses. First, the preponderance of blue beads throughout the task is likely to result in Type 1 processing outputting an intuition that the beads are being drawn from the mainly blue jar. Because this intuition comes to mind quickly and fluently, it should be associated with a "feeling of rightness" (Thompson, Prowse Turner, & Pennycook, 2011) that makes the intuitive response highly salient (Thompson et al., 2012). Consequently, if Type 2 processing does not intervene, a participant might stop requesting beads because they are overconfident about the strength of the evidence in support of their intuition. An alternative explanation is more closely linked to motivation than reasoning *per se*. Analytic cognitive style is associated with being motivated to exert effort so as to hold evidence-based beliefs (i.e., to be "actively open-minded"; Haran, Ritov, & Mellers, 2013; Toplak et al., 2011). Consequently, given the low costs associated with data gathering (i.e. a little additional time spent performing the task), participants with an analytic cognitive style could be motivated to persist with additional data gathering. Research that utilises data gathering paradigms that include normatively correct responses, monetary incentivisation, and control for risk aversion (e.g., van der Leer et al., 2015; van der Leer & McKay, 2014) could be used to tease apart these distinct (but not mutually exclusive) explanations for the association between analytic cognitive style and data gathering.

3.5.1 Limitations and directions for future research

The present study has a number of limitations. First, after controlling for other variables, the analytic cognitive measure accounted for only 4% of the variance in data gathering, and it is not clear that this should be interpreted as a causal connection. Future research could take an experimental approach to examine causality with more rigour. Previous research has shown that priming an analytic thinking style can increase religious belief (Gervais & Norenzayan, 2012; Shenhav, Rand, & Greene, 2012), and research using a similar paradigm could be used to investigate the effect of analytic thinking primes on both data gathering and delusional ideation.

Second, it is possible that the fact that data were collected using Mechanical Turk could have driven up Type 1 processing since participants might have been hurrying to complete the study to maximise the amount of money they make in a given period of time. This is, of course, possible. However, this problem is ubiquitous in psychology research. In a typical lab-based psychology study participants might likewise be hurrying to complete the task so that they can collect their payment and leave. We suggest that future research could benefit from using paradigms developed by behavioural economists in which participants are paid proportionate to how well they perform at data gathering tasks.

Third, in this study we did not collect self-report measures for the prior probabilities or posterior probabilities that participants assigned to the two jars. Consequently, we are unable to analyse the extent to which the probability judgments of participants conform to normative standards of belief updating derived by Bayes' Theorem. We suggest that it would be very worthwhile for future research to probe prior and posterior probability judgments so that the relationship between draws to decision, analytic thinking style, and delusional ideation can be investigated from a Bayesian perspective.

Fourth, a measure of subclinical delusional ideation, rather than clinical delusions, was examined. We suggest that there are at least two reasons why this might account for our failure to find an association with data gathering. First, although meta-analysis provides evidence for a negative association between delusional ideation and data gathering, the effect size appears to be very small (Ross et al., 2015). A number of review articles have suggested that associations are larger and more robust when considering clinical delusions (Garety &

Freeman, 1999; Garety et al., 2013); a claim that will be put to the test in an upcoming meta-analysis (Taylor, Hutton, & Dudley, 2014). Second, although there is evidence that clinical delusions lie on a continuum with normal belief (Bentall, 2003; Linscott & van Os, 2013; van Os, Hanssen, Bijl, & Ravelli, 2000; van Os, Linscott, Myin-Germeys, Delespaul, & Krabbendam, 2009), the continuum model has limitations (David, 2010; Lawrie, Hall, McIntosh, Owens, & Johnstone, 2010). With respect to the current study, it could be the case that the wording of the PDI items better tap a propensity to *entertain* delusion-like thoughts, rather than actually *endorse* delusion-like beliefs; data gathering may relate more strongly to the latter (a noteworthy feature of most of the PDI items is that they ask participants whether they have ever felt "as if" they have had an unusual experience or belief). This possibility finds tentative support from the present study since data gathering is associated more strongly with the Paranormal Belief Scale (which asks people *directly* about their beliefs without an "as if" qualification) than the PDI. Consequently, we suggest that an important direction for future research is to examine whether data gathering in the beads task discriminates between clinically deluded and control populations independently of analytic cognitive style.

3.6 Conclusion

We examined the extent to which delusional ideation and analytic cognitive style predict data gathering in the largest beads task study to date. We found that increased analytic cognitive style predicts greater data gathering, even when controlling for delusional ideation, cognitive ability, paranormal belief, and religiosity. Conversely, we did not find evidence for an association between delusional ideation and data gathering; a result that can be harmonised with the small effect size for this association found in a recent meta-analysis.

3.7 References

Aimola Davies, A. M., & Davies, M. (2009). Explaining pathologies of belief. In M. R. Broome & L. Bortolotti (Eds.), *Psychiatry as cognitive neuroscience: philosophical perspectives* (pp. 285-323). Oxford, UK: Oxford University Press

Barr, N., Pennycook, G., Stolz, J. A., & Fugelsang, J. A. (2015a). The brain in your pocket: evidence that smartphones are used to supplant thinking. *Computers in Human Behavior, 48*, 473-480. doi: 10.1016/j.chb.2015.02.029

Barr, N., Pennycook, G., Stolz, J. A., & Fugelsang, J. A. (2015b). Reasoned connections: a dual-process perspective on creative thought. *Thinking & Reasoning, 21*(1), 61-75. doi: 10.1080/13546783.2014.895915

Bentall, R. P. (2003). *Madness explained: psychosis and human nature*. London, UK: Penguin.

Brosnan, M., Hollinworth, M., Antoniadou, K., & Lewton, M. (2014). Is empathizing intuitive and systemizing deliberative? *Personality and Individual Differences, 66*, 39-43. doi: 10.1016/j.paid.2014.03.006

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3-5. doi: 10.1177/1745691610393980

Chandler, J., Mueller, P. A., & Paolacci, G. (2013). Nonnaïveté among Amazon Mechanical Turk workers: consequences and solutions for behavioral researchers. *Behavior Research Methods, 46*, 112-130. doi: 10.3758/s13428-013-0365-7

Cheyne, J. A., & Pennycook, G. (2013). Sleep paralysis postepisode distress: modeling potential effects of episode characteristics, general psychological distress, beliefs, and cognitive style. *Clinical Psychological Science, 1*(2), 135-148. doi: 10.1177/2167702612466656

Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research, 34*(3), 315-346. doi: 10.1207/S15327906MBR3403_2

Coltheart, M. (2007). The 33rd Sir Frederick Bartlett Lecture: cognitive neuropsychiatry and delusional belief. *The Quarterly Journal of Experimental Psychology, 60*(8), 1041-1062. doi: 10.1080/17470210701338071

Connors, M. H., & Halligan, P. W. (2015). A cognitive account of belief: a tentative roadmap. *Frontiers in Psychology, 5*(1588), 1-14. doi: 10.3389/fpsyg.2014.01588

Corlett, P. R., & Fletcher, P. C. (2014). Computational psychiatry: a Rosetta Stone linking the brain to mental illness. *Lancet Pscyhiatry, 1*(5), 399-402.

David, A. S. (2010). Why we need more debate on whether psychotic symptoms lie on a continuum with normality. *Psychological Medicine, 40*(12), 1935-1942. doi: 10.1017/S0033291710000188

Davies, J. A., & Smith, T. W. (1994). *General social surveys 1972-1994: cumulative codebook*. Chicago, Il: National Opinoin Research Center.

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: not always winning but at least taking part. *Cognition, 113*(1), 45-61. doi: 10.1016/j.cognition.2009.07.009

Evans, J. S. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning, 13*(4), 321-339. doi: 10.1080/13546780601008825

Evans, J. S. B. T. (2010). *Thinking twice: two minds in one brain*. Oxford, UK: Oxford University Press.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: advancing the debate. *Perspectives on Psychological Science, 8*(3), 223-241. doi: 10.1177/1745691612460685

Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics: and sex and drugs and rock 'n' roll* (4th ed.). Ca, USA: Sage.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25-42. doi: 10.1257/089533005775196732

Freeman, D., Evans, N., & Lister, R. (2012). Gut feelings, deliberative thought, and paranoid ideation: A study of experiential and rational reasoning. *Psychiatry Research, 197*(1-2), 119-122. doi: 10.1016/j.psychres.2011.12.031

Freeman, D., & Garety, P. A. (2014). Advances in understanding and treating persecutory delusions: a review. *Social Psychiatry and Psychiatric Epidemiology, 49*(8), 1179-1189. doi: 10.1007/s00127-014-0928-7

Freeman, D., Lister, R., & Evans, N. (2014). The use of intuitive and analytic reasoning styles by patients with persecutory delusions. *Journal of Behavioral Therapy and Experimental Psychiatry, 45*(4), 454-458. doi: 10.1016/j.jbtep.2014.06.005

Garety, P. A., & Freeman, D. (1999). Cognitive approaches to delusions: a critical review of theories and evidence. *British Journal of Clinical Psychology, 38*, 113-154. doi: 10.1348/014466599162700

Garety, P. A., & Freeman, D. (2013). The past and future of delusions research: from the inexplicable to the treatable. *British Journal of Psychiatry, 203*(5), 327-333. doi: 10.1192/bjp.bp.113.126953

Garety, P. A., Freeman, D., Jolley, S., Dunn, G., Bebbington, P. E., Fowler, D. G., . . . Dudley, R. (2005). Reasoning, emotions, and delusional conviction in psychosis. *Journal of Abnormal Psychology, 114*(3), 373-384. doi: 10.1037/0021-843X.114.3.373

Garety, P. A., Freeman, D., Jolley, S., Ross, K., Waller, H., & Dunn, G. (2011). Jumping to conclusions: the psychology of delusional reasoning. *Advances in Psychiatric Treatment, 17*(5), 332-339. doi: 10.1192/apt.bp.109.007104

Garety, P. A., Joyce, E., Jolley, S., Emsley, R., Waller, H., Kuipers, E., . . . Freeman, D. (2013). Neuropsychological functioning and jumping to conclusions in delusions. *Schizophrenia Research, 150*, 570-574. doi: 10.1016/j.schres.2013.08.035

Garety, P. A., Kuipers, E., Fowler, D., Freeman, D., & Bebbington, P. E. (2001). A cognitive model of the positive symptoms of psychosis. *Psychological Medicine, 31*, 189-195. doi: 10.1017/s0033291701003312

Garety, P. A., Waller, H., Emsley, R., Jolley, S., Kuipers, E., Bebbington, P., . . . Freeman, D. (2015). Cognitive mechanisms of change in delusions: an experimental investigation targeting reasoning to effect change in paranoia. *Schizophrenia Bulletin, 41*(2), 400-410. doi: 10.1093/schbul/sbu103

Gervais, W. M., & Norenzayan, A. (2012). Analytic thinking promotes religious disbelief. *Science, 336*(6080), 493-496. doi: 10.1126/science.1215647

Haran, U., Ritov, I., & Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgment and Decision Making, 8*(3), 188-201. doi: 10.1017/cbo9780511840265.012

Huang, M. H., & Hauser, R. M. (1998). Trends in Black-White test score differentials: II The WORDSUM vocabulary test. In U. Neisser (Ed.), *The rising curve: long-term gains in IQ and related measures* (pp. 303-332). Washington, DC: American Psychological Association.

Huq, S. F., Garety, P. A., & Hemsley, D. R. (1988). Probabilistic judgements in deluded and non-deluded subjects. *Quarterly Journal of Experimental Psychology Section A, 40*(4), 801-812. doi: 10.1080/14640748808402300

Irwin, H. J., Drinkwater, K., & Dagnall, N. (2014). Are believers in the paranormal inclined to jump to conclusions? *Australian Journal of Parapsychology, 14*(1), 69-82.

Kahneman, D. (2003). Maps of bounded rationality: psychology for behavioral economics. *The American Economic Review, 93*(5), 1449-1475. doi: 10.1257/000282803322655392

Kahneman, D. (2011). *Thinking, fast and slow*: Farrar, Straus and Giroux.

Lawrence, E., & Peters, E. R. (2004). Reasoning in Believers in the Paranormal. *The Journal of Nervous and Mental Disease, 192*(11), 727-733. doi: 10.1097/01.nmd.0000144691.22135.d0

Lawrie, S. M., Hall, J., McIntosh, A. M., Owens, D. G. C., & Johnstone, E. C. (2010). The 'continuum of psychosis': scientifically unproven and clinically impractical. *The British Journal of Psychiatry, 197*(6), 423-425. doi: 10.1192/bjp.bp.109.072827

Linscott, R. J., & van Os, J. (2013). An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: On the pathway from proneness to persistence to dimensional expression across mental disorders. *Psychological Medicine, 43*(6), 1133-1149. doi: 10.1017/S0033291712001626

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making, 21*(1), 37-44. doi: 10.1177/0272989x0102100105

Maher, B. A. (2004). Delusion. In H. E. Adams & P. B. Sutker (Eds.), *Comprehensive handbook of psychopathology* (3rd ed., pp. 309-340). New York, NY: Springer.

Malhotra, A. K., Krosnick, J. A., & Haertel, E. (2007). The psychometric properties of the GSS Wordsum Vocabulary Test (GSS Methodological Report No. 111). Chicago: NORC.

Moritz, S., Andreou, C., Schneider, B. C., Wittekind, C. E., Menon, M., Balzan, R. P., & Woodward, T. S. (2014). Sowing the seeds of doubt: a narrative review on metacognitive training in schizophrenia. *Clinical Psychology Review, 34*(4), 358-366. doi: 10.1016/j.cpr.2014.04.004

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*(4), 867-872. doi: 10.1016/j.jesp.2009.03.009

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2013). Cognitive style and religiosity: the role of conflict detection. *Memory & Cognition, 42*, 1-10. doi: 10.3758/s13421-013-0340-7

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2014). The role of analytic thinking in moral judgements and values. *Thinking & Reasoning, 20*(2), 188-214. doi: 10.1080/13546783.2013.865000

Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2013). Belief bias during reasoning among religious believers and skeptics. *Psychonomic Bulletin & Review, 20*(4), 806-811. doi: 10.3758/s13423-013-0394-3

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition, 123*(3), 335-346. doi: 10.1016/j.cognition.2012.03.003

Peters, E. R., Day, S., McKenna, J., & Orbach, G. (1999). Delusional ideation in religious and psychotic populations. *British Journal of Clinical Psychology, 38*, 83-96. doi: 10.1348/014466599162683

Peters, E. R., Joseph, S. A., Day, S., & Garety, P. A. (2004). Measuring delusional ideation: The 21-Item Peters et al. Delusions Inventory (PDI) *Schizophrenia Bulletin, 30*(4), 1005-1022. doi: 10.1093/oxfordjournals.schbul.a007116

Peters, E. R., Joseph, S. A., & Garety, P. A. (1999). Measurement of delusional ideation in the normal population: introducing the PDI (Peters et al. Delusions Inventory). *Schizophrenia Bulletin, 25*(3), 553-576. doi: 10.1093/oxfordjournals.schbul.a033401

Phillips, L. D., & Edwards, W. (1966). Conservatism in simple probabilistic inference. *Journal of Experimental Psychology, 72,* 346-354.

Ross, R. M., McKay, R., Coltheart, M., & Langdon, R. (2015). Jumping to conclusions about the Beads Task? A meta-analysis of delusional ideation and data-gathering. *Schizophrenia Bulletin, Advance online publication*. doi: 10.1093/schbul/sbu187

Schwartz, L. M., Woloshin, S., Black, W., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annuals of Internal Medicine, 127*(11), 966-972.

Shenhav, A., Rand, D. G., & Greene, J. D. (2012). Divine intuition: cognitive style influences belief in God. *Journal of Experimental Psychology: General, 141*(3), 423-428. doi: 10.1037/a0025391

Speechley, W. J., Murray, C. B., McKay, R. M., Munz, M. T., & Ngan, E. T. C. (2010). A failure of conflict to modulate dual-stream processing may underlie the formation and maintenance of delusions. *European Psychiatry, 25*(2), 80-86. doi: 10.1016/j.eurpsy.2009.05.012

Speechley, W. J., & Ngan, E. T. C. (2008). Dual-stream modulation failure: a novel hypothesis for the formation and maintenance of delusions in schizophrenia. *Medical Hypotheses, 70*(6), 1210-1214. doi: 10.1016/j.mehy.2007.11.017

Speechley, W. J., Whitman, J. C., & Woodward, T. S. (2010). The contribution of hypersalience to the "jumping to conclusions" bias associated with delusions in schizophrenia. *Journal of Psychiatry & Neuroscience, 35*(1), 7-17. doi: 10.1503/jpn.090025

Speechley, W. J., Woodward, T. S., & Ngan, E. T. (2013). Failure of conflict to modulate central executive network activity associated with delusions in schizophrenia. *Frontiers in Psychiatry, 4*, 113. doi: 10.3389/fpsyt.2013.00113

Stanovich, K. E. (2011). *Rationality and the reflective mind*: Oxford, UK: Oxford University Press.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*(4), 672-695. doi: 10.1037/0022-3514.94.4.672

Taylor, P., Hutton, P., & Dudley, R. (2014). Rationale and protocol for a systematic review and meta-analysis on reduced data gathering in people with delusions. *Systematic Reviews, 3*(1), 44. doi: 10.1186/2046-4053-3-44

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63*(3), 107-140. doi: 10.1016/j.cogpsych.2011.06.001

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. A. (2012). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128*, 237-251. doi: 10.1016/j.cognition.2012.09.012

Tobacyk, J. J. (2004). A revised paranormal belief scale. *The International Journal of Transpersonal Studies, 23*, 94-98. doi: 10.1037/t14015-000

Tobacyk, J. J., & Milford, G. (1983). Belief in paranormal phenomena: assessment instrument development and implications for personality functioning. *Journal of Personality and Social Psychology, 44*(5), 1029-1037. doi: 10.1037//0022-3514.44.5.1029

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39*(7), 1275-1289. doi: 10.3758/s13421-011-0104-1

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Assessing miserly information processing: an expansion of the Cognitive Reflection Test. *Thinking & Reasoning, 20*(2), 147-168. doi: 10.1080/13546783.2013.844729

Trippas, D., Pennycook, G., Verde, M. F., & Handley, S. J. (in press). Better but still biased: the link between analytic cognitive style and belief bias. *Thinking & Reasoning. Advance online publication*.

van der Leer, L., Hartig, B., Goldmanis, M., & McKay, R. (2015). Delusion-proneness and 'jumping to conclusions': relative and absolute effects. *Psychological Medicine, 45*(6), 1253-1262. doi: 10.1017/s0033291714002359

van der Leer, L., & McKay, R. (2014). "Jumping to conclusions" in delusion-prone participants: an experimental economics approach. *Cognitive Neuropsychiatry, 19*(3), 257-267. doi: 10.1080/13546805.2013.861350

van Os, J., Hanssen, M., Bijl, R. V., & Ravelli, A. (2000). Strauss (1969) revisited: a psychosis continuum in the general population? *Schizophrenia Research, 45*, 11-20. doi: 10.1016/s0920-9964(00)90323-2

van Os, J., Linscott, R. J., Myin-Germeys, I., Delespaul, P., & Krabbendam, L. (2009). A systematic review and meta-analysis of the psychosis continuum: Evidence for a psychosis proneness–persistence–impairment model of psychotic disorder. *Psychological Medicine, 39*(02), 179-195. doi: 10.1017/S0033291708003814

Chapter Four: Retrospective revaluation and the prediction error theory of delusions

4.1 Abstract

It has recently been proposed that delusions are caused by aberrant prediction error signalling. Key evidence for this theory comes from a series of neuroimaging studies that adapted the “allergist task” from the causal learning literature to investigate the relationship between “retrospective revaluation” of learned associations and neural signatures of “prediction error” in healthy and delusion-prone populations. We argue that an important limitation of these studies is that the behavioural evidence that they reported to show that retrospective revaluation had occurred is not strong. To address this limitation we developed a closely matched behavioural experiment that tests for retrospective revaluation—both “unovershadowing” and “backward blocking”—more rigorously. Our results provided clear evidence for backward blocking, but no evidence for unovershadowing. Given that the studies that linked aberrant prediction error signalling to delusions or delusion-proneness focus exclusively on the unovershadowing trials, and did not analyse the backward blocking trials, we argue that our results raise doubts about the empirical evidence for a putative link between aberrant prediction error signalling and delusions.

4.2 Introduction

Delusions can be roughly described as false beliefs that are characterised by impossibility, incorrigibility, and unwarranted subjective certainty (e.g., Coltheart, Langdon, & McKay, 2011). Despite serious impasses in defining delusions more precisely (David, 1999), research on the causes of delusions is urgent because they frequently manifest as highly distressing symptoms associated with a variety of disorders, including schizophrenia, dementia, and traumatic brain injury.

Research on the cognitive and the neurobiological underpinnings of delusions has tended to progress independently. Cognitive psychologists (e.g., Coltheart et al., 2011) have tended to focus on developing cognitive models that aim to explain the content and persistence of delusions, while ignoring the neurobiology of what goes on in the brain. By contrast, pharmacologists and neurobiologists (e.g., Howes & Kapur, 2009) have tended to focus on identifying neurobiological abnormalities and testing the efficacy of different pharmacological interventions, while setting aside issues pertaining to rationality and information processing. Consequently, limited progress has been made in bridging the cognitive/neurobiological divide despite the fact that a good case can be made that explanations of mental illness need to "combine references to brute, a-rational neural mechanisms and to the rationality of persons" (Graham, 2010, p. 7).

Recently it has been argued that developments in cognitive neuroscience that focus on the role that prediction plays in brain functioning may hold the key to integrating the cognitive and neurobiological levels of analysis of the mind/brain (Clark, 2013; Friston & Stephan, 2007; Hohwy, 2013). According to the "predictive brain hypothesis," the brain is essentially a "hierarchical prediction machine" that "support[s] perception and action by constantly attempting to match incoming sensory inputs with top-down expectations or predictions" (Clark, 2013, p. 181). In brief, sensory information from the external world is perceived and interpreted with reference to an agent's predictions about what sensations it is likely to experience (a top-down process); and, concurrently, the incoming sensory information is used to update the agent's model of the world (a bottom-up process) and thereby influence future predictions of expected sensations. Of particular importance in this model is that "prediction error" is used as the bidirectional signal between adjacent layers of representations: it signals an unexpected sensory event (that is likely to be important and thus should be attended to), and it also signals that the predictive model is in error (and thus may need to be updated).

It has recently been proposed that delusions and other symptoms of psychosis are related to aberrant prediction error signalling (Corlett, Frith, & Fletcher, 2009; Corlett, Krystal, Taylor, & Fletcher, 2009; Corlett et al., 2006, 2007; Corlett, Honey, Krystal, & Fletcher, 2011; Corlett, Taylor, Wang, Fletcher, & Krystal, 2010; Corlett, Simons, et al., 2009; Fletcher & Frith, 2009). More specifically, delusions are hypothesised to be caused by abnormal neural processing (possibly linked to dopaminergic dysfunction) that result in "inappropriate mismatches between expectancy and experience [that] engender[s] prediction error where there ought to be none, driving new and aberrant learning directly and through the allocation of attention towards irrelevant but potentially explanatory cues" (Corlett et al., 2010, p. 357). This proposed mismatch between expectancy and experience is consistent with reports from people in the early stages of psychosis who frequently describe the world as seeming different and unexplainable, with undue salience being inappropriately linked to innocuous features of the environment (Fletcher & Frith, 2009).

To date, the primary empirical evidence linking aberrant prediction error signalling to delusions comes from a series of three brain imaging studies of prediction error-mediated learning (Corlett et al., 2004, 2006, 2007). These studies adapted the behavioural "allergist task" (Dickinson & Burke, 1996) for use in a functional magnetic resonance imaging (fMRI) brain scanner. In these studies, participants were asked to imagine that they are allergists who are studying food allergies in an imaginary client. They are shown a sequence of one- and two-food meals that are being fed to their client; their job is to learn to predict which meals are associated with an allergic response and which are not. Table 4.1 shows how food cues and allergic responses are combined in the three stages of the allergist task used in the studies by Corlett and colleagues.

Stage 1: Training (12)	Stage 2: Retrospective revaluation (6)	Stage 3: Violation or confirmation of learned expectancies (6)	Role
A_1B_1+	A_1+	B_1+	Backward blocking of B in Stage 2 [& violation of expectancy about B in Stage 3]
A_2B_2+	A_2+	B_2-	Backward blocking of B in Stage 2 [& confirmation of expectancy about B in Stage 3]
C_1D_1+	C_1-	D_1+	Unovershadowing of D in Stage 2 [& confirmation of expectancy about D in Stage 3]
C_2D_2+	C_2-	D_2-	Unovershadowing of D in Stage 2 [& violation of expectancy about D in Stage 3]
E_1F_1-	F_1-	F_1-	[Control for neuroimaging data]
E_2F_2-	F_2-	F_2-	[Control for neuroimaging data]
I+	I+	I+	Control for backward blocking of B in Corlett *et al.* (2007)
J-	J-	J-	Control for unovershadowing of D in Corlett *et al.* (2007)

Table 4.1 Experimental design of Corlett *et al.*'s (2004, 2006, 2007) allergist task.

Note. Foods are represented with either a letter (A-J) or a letter paired with a numeric subscript (Corlett and colleagues use these subscripts to indicate that two different cues are treated identically during Stages 1 and 2). Feedback indicating an allergic response is represented with a "+", and feedback indicating no allergic response is represented with a "-". Numbers in brackets indicate number of repetitions of each trial-type at each stage. Roles for cues presented in square brackets pertain to neuroimaging data and are not discussed further because this paper focuses on evaluating the behavioural data.

During Stage 1, meals were presented to the imaginary client. Some meals are comprised of one food, while others are comprised of two foods. At each meal presentation participants had to predict whether or not the meal would cause an allergic response. Predictions were collected using a two-button control box in an fMRI brain scanner. Participants were asked to hold down one button to predict an allergic response, and the other button to predict the absence of an allergic response. Furthermore, participants were instructed to hold down the button for an amount of time proportional to their degree of confidence in their judgement. Immediately after making a prediction, participants were given feedback indicating whether or not an allergic response actually occurred. Initially they had to guess whether or not a response would occur, but after repeated presentations and feedback they came to make correct predictions. Each trial type was repeated 12 times, with the order of presentation being random, before moving onto Stage 2.

During Stage 2, all two-food meals from Stage 1 (i.e., A_1B_1+, A_2B_2+, C_1D_1+, C_2D_2+, E_1F_1-, E_2F_2-) were broken up so that one of the foods from each meal was presented alone, and the other food was not presented at all. The purpose of splitting these two-food meals was to study "retrospective revaluation", which is defined as an indirect change in a cue's associative strength resulting from later information (Shanks, 2010). Some of these split up one-food meals were paired with an allergic response – the aim being to induce "backward blocking" (where presentation of an A+ cue weakens a previously learned association between B and an allergic response during AB+ training). And some of these one-food meals were paired with the absence of an allergic response – the aim being to induce "unovershadowing" (where presentation of a C- cue strengthens a previous learned association between D and an allergic response during the CD+ training). Each trial type was repeated 6 times, with the order of presentation being random, before moving onto Stage 3.

During Stage 3, all foods were presented alone as one-food meals. The crucial trials for testing for retrospective revaluation were those trials that presented food cues that had appeared in two-food meals paired with an allergic response during Stage 1 but had not been presented during Stage 2 (i.e., B_1+, B_2-, D_1+, and D_2-). Furthermore, for the purposes of correlating cue with neural signatures of "surprise" (i.e. prediction error), the authors classified B_2- and D_1+ trials during Stage 3 as showing "confirmation of learned expectancy," and B_1+ and D_2- trails during Stage 3 as showing "violation of learned expectancy". The authors predicted that violation trials would evoke more prediction error than confirmation trials. They indexed prediction error as the magnitude of change in hemodynamic responses in a specific area of right lateral prefrontal cortex (RLPFC) that had been implicated in earlier studies by the same research group as being involved in prediction error signalling (Fletcher et al., 2001; Turner et al., 2004). Each trial type was repeated 6 times, with the order of presentation being random. When Stage 3 was completed the experiment came to an end.

Although each of the three studies by Corlett *et al.* (2004, 2006, 2007) used the allergist task described above, the statistical comparisons used to test hypotheses were different in each study. Unfortunately, the behavioural evidence for retrospective revaluation provided by each of these studies is not strong (Griffiths, Langdon, Le Pelley, Coltheart, 2014). In their first study, Corlett *et al.* (2004) examined prediction error mediated learning in healthy adults. They argued that their results provide evidence that activation in RLPFC is associated with normal prediction error mediated retrospective revaluation. The test for behavioural evidence of retrospective revaluation that they reported in this study was a comparison of the strength of participants' expectations about the first trial of cue B during Stage 3 (which had been through backward blocking trials in Stage 2) to the first presentation of cue D during Stage 3 (which had been through unovershadowing trials in Stage 2). Corlett *et al.* (2004) hypothesised that if retrospective evaluation had occurred then "the unovershadowed items should be accompanied by a stronger initial prediction of an allergic response than backward blocked items" (p. 880). We agree that this is a useful statistical comparison for testing for evidence of an overall effect of retrospective revaluation.

However, the evidence for the predicted difference that they reported is not strong: a one tailed *t* test yielding a non-significant test statistic ($p = .08$), which the authors interpreted as "a trend for a difference" (p. 880). Moreover, comparing cue B versus cue D at the beginning of Stage 3 is not informative with respect to whether or not this putative trend for a difference was the result of backward blocking, or unovershadowing, or both.

Before discussing the details of the two follow up studies, it is important to draw attention to a general point: in neither of the follow up studies did Corlett and colleagues report the comparison of cue B versus cue D during Stage 3 that they reported in their first study; rather, they focused exclusively on statistical analysis of cues that had been through unovershadowing training. Their justification for not reporting a test for an overall effect of retrospective revaluation or a test for backward blocking in their follow up studies was that earlier research (Larkin, Aitken & Dickinson 1998) provided evidence that unovershadowing induces a larger change in causal rating than backward blocking. However, this justification is not strong—a fuller reading of the literature shows that earlier research is not consistent on this point. Although the study by Larkin *et al.* (1998) provides evidence for a larger unovershadowing effect, it is the only study we could identify that does so. By contrast, we have identified two studies that report a larger backward blocking effect (De Houwer & Beckers, 2002; Wasserman, Kao, Van Hamme, Katagiri, & Young, 1996), and two studies that report backward blocking and unovershadowing effects of comparable magnitude (Le Pelley & McLaren, 2001; Wasserman & Berglan, 1998). Given this mixed evidence we suggest that it is important to analyse both unovershadowing and backward blocking data to get a full and accurate picture of retrospective revaluation in the allergist task.

Turning to the second study, Corlett *et al.* (2006) examined the effect of the drug ketamine, which can produce psychotic-like phenomenology in healthy people, on prediction error mediated retrospective revaluation. Using a within-subjects placebo-controlled design they ran participants through their allergist task during two separate testing sessions: low-dose ketamine and placebo. (They also ran their participants through a high-dose ketamine testing session, but they did not report the results from this testing session because "performance was inadequate" p. 612.) They argued that their results provide evidence that low-dose ketamine perturbs prediction error mediated learning activity in RLPFC. However, again, we note that the behavioural evidence that retrospective revaluation had actually occurred is not strong. In this paper, the authors plotted the mean predictive response for the putatively unovershadowed cue (i.e., cue D) during its first presentation at Stage 3 to index participants' expectations for the placebo condition compared to the ketamine condition (see Corlett et al. 2006; Figure 3c, p. 615). They stated that "this [responses to the first presentation of cue D at Stage 3 shown in the figure] can be taken as an index of the extent to which [participants] have indeed revalued the items that were absent at stage 2" (p. 615). In addition, they report that the putatively unovershadowed cue D had a higher mean casual rating in the placebo condition than the ketamine condition. This mean difference did not reach statistical significance (p = .09), but the authors interpret this as evidence for a "trend towards a difference" (p. 615) which they suggest shows "a reduction in the magnitude of unovershadowing" (p. 615) in the ketamine condition. However, any claim that this plot and statistical comparison can be used to provide evidence for unovershadowing is problematic because they merely compare the placebo and ketamine testing sessions with respect to expectations for cue D *at the beginning of Stage 3*. There are many potential explanations for differences between differences in causal ratings at Stage 3 in different testing sessions that are not the direct result of unovershadowings—for example, it may be that participants simply give lower confidence ratings to this cue when they are under the influence of the psychoactive effects of ketamine compared to a placebo. To provide direct evidence that unovershadowing of cue D had occurred during Stage 2 training, it is important to provide evidence that there is an increase in the rating of cue D at the beginning of Stage 3 (i.e. after the Stage 2 training) *relative to the rating of cue D after the Stage 1 training*. No such evidence is reported.

In their third study, Corlett *et al.* (2007) compared retrospective revaluation in participants diagnosed with first-episode psychosis to healthy controls, to study prediction error mediated learning in psychosis. They argued that their psychosis patients and their healthy controls showed different patterns of activation in RLPFC during prediction error mediated retrospective revaluation. However, once again, we note that the behavioural evidence for retrospective revaluation is not strong. In this study, the authors found that the first presentation at Stage 3 of cue D (which had been through unovershadowing training) showed a significantly higher causal rating than cue J (which had never been paired with an allergic response) [$F(1,22) = 84.37$, $p < .001$]. They claimed that this provides evidence that unovershadowing occurred. However, unfortunately, this comparison is not an appropriate test for unovershadowing of cue D because it does not control for simple associative learning. Given that cue J was presented during Stage 1 and Stage 2 without an allergic response, associative learning theory predicts that no association between cue J and an allergic response will form, and a low causal rating for cue J is therefore to be expected at the first presentation of cue J at Stage 3. By contrast, although cue D was not presented during Stage 2, it was paired with an allergic response during Stage 1 (as CD+) so associative learning theory predicts that an association between cue D and an allergic response is likely to form via simple associative learning during Stage 1 *even in the complete absence of unovershadowing during Stage 2*. As such, the statistically significant difference in causal ratings between cue D and cue J at Stage 3 does not provide strong evidence that unovershadowing of cue D occurred during Stage 2.

To summarise, we have presented serious concerns about the behavioural evidence for retrospective revaluation in each of these three studies. In the first study, Corlett *et al.* (2004) used an appropriate comparison between cues to test for an overall effect of retrospective revaluation, but found only suggestive statistical evidence (one-tailed *t* test: $p = .08$). Furthermore, although they argued that this putative “trend towards a difference” (Corlett et al., 2004, p. 615) is more likely to be the result of unovershadowing than backwards blocking, we found that when the literature is surveyed more widely their argument loses its force. In the second study, Corlett *et al.* (2006) claimed that they found evidence for unovershadowing (they did not test for backward blocking). However, we argued that their plot and analysis of the cue that had been through unovershadowing training did not provide evidence that the putative change in the cue’s causal rating was the result of unovershadowing training and thus should not have been interpreted as providing evidence for unovershadowing. In the third study, Corlett *et al.* (2007) claimed that they found evidence for unovershadowing (they did not test for backward blocking). However, we argued that their analysis did not control for simple associative learning and thus should not have been interpreted as providing evidence for unovershadowing.

In the present study, we develop a behavioural allergist task that is closely matched to the neuroimaging allergist task used in Corlett and colleagues’ studies—with one crucial advantage: our experimental design includes testing phases between learning blocks, which enable us to test for behavioural evidence of backward blocking and unovershadowing independently of each other while controling for simple associative learning. Because our study focuses exclusively on testing for behavioural evidence of retrospective revaluation (i.e., we do not collect any neuroimaging data), we do not evaluate the acquisition or analysis of neuroimaging data presented by Corlett and colleagues in their studies. Nevertheless, it is worth noting that there exists considerable debate about how (and if) imaging data should be integrated with behavioural data to test hypotheses about cognitive processes (Shallice & Cooper, 2011; also see special issue introduced by Mather, Cacioppo, & Kanwisher, 2013).

4.3 Methods

38 naïve students and staff from Macquarie University (13 male, 25 female; mean age = 27.3 years, SD = 3.5 years) participated in the study in return for AU$10. Testing was conducted in a 4-booth computer laboratory. LiveCode (University Edition, Version 5.5.1) was used to present all instructions and experimental trials, and record predictive responses and causal ratings via a computer mouse.

Participants were seated at individual computers and asked to follow the onscreen instructions. The first screens consisted of instructions explaining that they should imagine themselves to be scientists who study food allergies and would be confronted with a subject in their study, "Mr. X," who suffers an allergic response after eating some foods, and that their task is to use feedback to learn to make accurate predictions about which particular meals cause an allergic response. The experiment had five key phases that were presented in the same order for all participants, as summarised in Table 4.2.

Table 4.2 Experimental design of our allergist task.

Test 1: Pre training	Stage 1: Training	Test 2: Post training	Stage 2: Induction of retrospective revaluation	Test 3: Post retrospective revaluation	Role
A, B	AB+	A, B	A+	A, B	Backward blocking of B
C, D	CD+	C, D	C-	C, D	Unovershadowing of D
E, F	EF-	E, F	E-	E, F	Filler
G, H	GH-	G, H	G-	G, H	Filler
I	I+	I	I+	I	Filler
J	J-	J	J-	J	Filler & "control" for unovershadowing of D
K	K+	K	K+	K	Filler
L	L-	L	L-	L	Filler

4.3.1 Test 1: Pre training

The question "If Mr. X ate this food, how likely is it that he would have an allergic reaction?" appeared on the screen, below which was presented the name of one food and a sliding scale marked from "0" ("He would definitely NOT have a reaction") to "50" ("Not sure") to "100" ("He would DEFINITELY have a reaction"). 12 food names were used: *beans, beef, corn, eggs, figs, mushrooms, olives, peaches, potatoes, rice, spinach, tomatoes*. For each participant, food names were randomly allocated to 12 cue codes: *A, B, C, D, E, F, G, H, I, J, K, L*. Participants made their ratings using a computer mouse to select a point on the sliding scale. Each of the foods was presented by itself once, with the order of presentation being random. Once all 12 foods had been evaluated participants moved onto Stage 1.

Test 1 was designed to measure the baseline expectancies of food-allergies prior to learning.

4.3.2 Stage 1: Training

For each trial, "Today Mr. X eats," appeared on the screen, below which was presented the names of one or two foods and the question "Will Mr. X have a reaction?" Participants used the computer mouse to make a forced choice, either "allergic reaction" or "no reaction". Immediately after participants responded, one of two outcomes appeared as feedback: either "no reaction" in green letters, bounded by a green rectangle; or "allergic reaction" in red letters, bounded by a red rectangle. Foods were presented in blocks comprising of 12 trials, with each food being presented once in random order. After the third block, participants' predictions were subject to a performance criterion: if they correctly predicted all food-outcome relationships during a block then Stage 1 ended and they moved onto Test 2. However, if they made at least one error then they were presented with another block of 12 trials, followed by another performance criterion, and so on, until they passed the performance criterion.

Stage 1 was designed to set up experimentally-induced expectancies. The key cues were two pairs of foods for which participants learned to expect a positive cue-outcome relationship (AB+, CD+). That is, participants were taught to expect that each of these food pairs always predicts an allergic response. Filler trials consisted of pairs of foods (EF-, GH-) and single foods (I+, J+, K-, L-) which were designed to ensure that participants must attend to the particular food items presented to predict the allergic outcome, as not all pairs of foods or single foods were associated with an allergic response or no response (see Table 4.2).

4.3.3 Test 2: Post training

Test 2 was identical to Test 1. Once all 12 foods had been evaluated participants moved onto Stage 2.

Test 2 was designed to measure the expectancies of food allergies learned during training in Stage 1. The key cues were two pairs of foods for which participants learned to expect a positive cue-outcome relationship during Stage 1 (AB+, CD+) (see Table 4.2).

4.3.4 Stage 2: Retrospective revaluation

Stage 2 was identical to Stage 1, with one crucial difference: all food pairs from Stage 1 were split up so that one of the foods from the initial pair was presented (A, C, E, G) and the other was not (B, D, F, H). Once participants passed the performance criterion they moved onto Test 3.

Stage 2 was designed to cause retrospective revaluation of the foods B and D that had been paired with foods A and C respectively and an allergic response during Stage 1. A+ was expected to cause backward blocking of B; and C- was expected to cause unovershadowing of D (see Table 4.2).

4.3.5 Test 3: Post retrospective revaluation

Test 3 was identical to Test 1 and Test 2. Once all 12 foods had been evaluated participants moved onto a debriefing screen and the experiment ended.

Test 3 was designed to measure the expectancies learned during Stage 2 and Stage 3. The key trials were for foods B and D because they were predicted to show backward blocking and unovershadowing respectively (see Table 4.2).

Our allergist task was explicitly designed to closely match the task developed by Corlett and colleagues, but with one crucial addition: test phases between the learning phases. Corlett and colleagues tested for behavioural evidence of retrospective revaluation using the first predictive responses to each cue at the beginning of Stage 3, and they measured the strength of associations by recording how long participants held down a button predicting an allergic response or another button predicting no allergic response. By contrast, our experimental design included three test phases (Test 1, Test 2, and Test 3) before and after each learning stage (Stage 1, and Stage 2) with responses during Test 2 and Test 3 being used to test for behavioural evidence of retrospective revaluation, and we measured the strength of associations by recording the rating that participants chose on a sliding scale from 0 to 100. The important advantage of our design is that it enables us to test for backward blocking and unovershadowing independently of each other.

4.4 Results

Figure 4.1 shows mean causal ratings across Test 1, 2, and 3 as a function of cue (A-L). Figure 4.1(a) shows "filler cues" (E, F, G, H, I, J, K, L). As expected, those cues that were consistently paired with an allergic response (I, J) have high mean causal ratings at both Test 2 and Test 3, and those consistently paired with no response (E, F, G, H, K, L) have low mean causal ratings at both Test 2 and Test 3. Figure 4.1(b) shows the cues we used for testing hypotheses about retrospective revaluation (B, D) and the cues that they were paired with during Stage 1 (A, C). The mean causal rating for cue B—which underwent AB+ training during Stage 1, and A+ backward blocking training during Stage 2—shows a somewhat lower mean causal rating at Test 3 than Test 2. This is consistent with backward blocking having occurred (backward blocking weakens the associative strength between a cue and an outcome). By contrast, the mean causal rating for cue D—which underwent CD+ training during Stage 1 and C- unovershadowing during Stage 2—shows a somewhat lower mean causal rating at Test 3 than Test 2. This is not consistent with unovershadowing having occurred (unovershadowing strengthens the associative strength between a cue and an outcome).

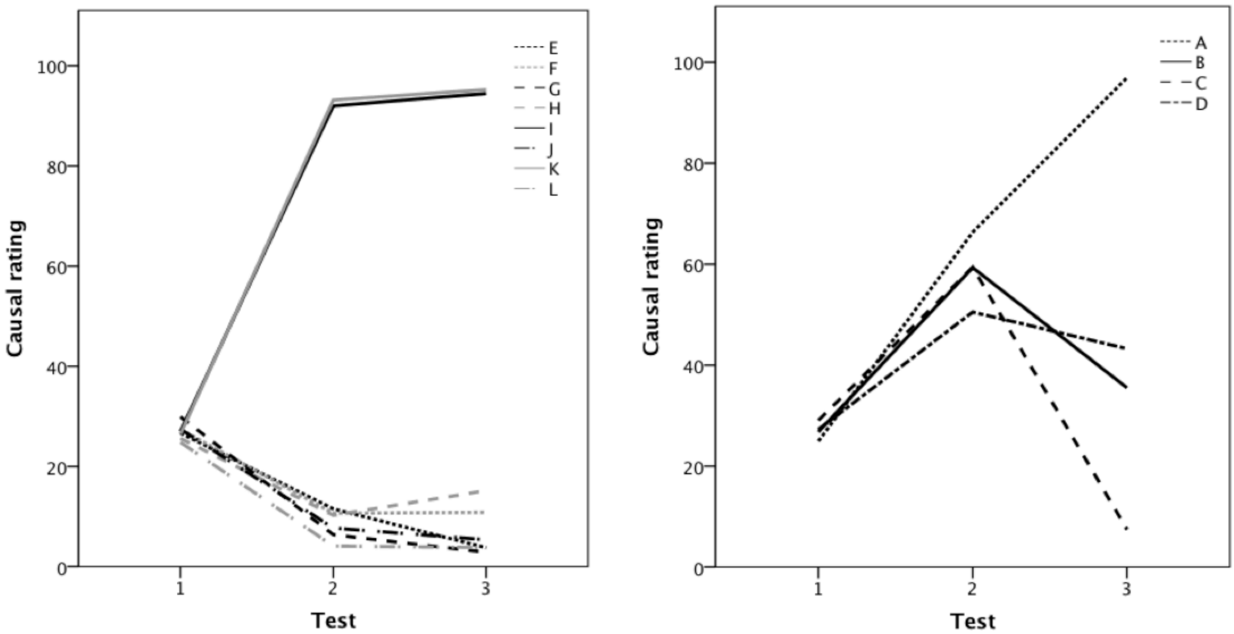


Figure 4.1 Mean casual ratings across Test 1, 2, and 3 as a function of cue (A-L). (a) The "filler cues" (E, F, G, H, I, J, K, L). (b) The cues used for hypothesis testing (B, D) and the cues that they are paired with during Stage 1 (A, C).

To test for an overall effect of retrospective revaluation we first considered the interaction between causal ratings at Test 2 and Test 3 for cue B (which had been through the backward blocking training) and cue D (which had been through the unovershadowing training). To do this we used a repeated measures general linear model with Test 2 (Post-

Training) versus Test 3 (Post Retrospective Revaluation) as one within-subjects factor, and cue B (which is hypothesised to show backward blocking—i.e. reduced ratings from Test 2 to Test 3) versus cue D (which is hypothesised to show unovershadowing—i.e. increased ratings from Test 2 to Test 3) as another within-subject factor. This analysis revealed no significant effect of cue [$F(1,37) = .07$, $p = .936$], a significant effect of Test [$F(1,37) = 13.058$, $p = .001$], and a significant cue by Test interaction [$F(1,37) = 4.809$, $p = .035$]. The cue by Test interaction provides evidence that retrospective revaluation had occurred during Stage 2 learning, but the interaction does not tell us what form of retrospective revaluation (unovershadowing, or backward block, or both) occurred.

Next we considered planned comparisons that we designed to replicate comparisons used by Corlett and colleagues. Our first replication followed Corlett *et al.* (2004) in testing for an overall effect of retrospective revaluation by comparing the Test 3 causal ratings for cue B and cue D. The mean causal rating for cue D (M = 43.26, SD = 39.15) was higher than for cue B (M = 35.53, SD = 34.47), but the mean difference was not statistically significant [$t(37) = 1.042$, $p = .304$]. The directionality of this difference is consistent with retrospective revaluation having occurred (because cue D went through unovershadowing training and cue B went through backward blocking training), but because the test statistic is far from being statistically significant it does not provide any positive evidence. Empirically, this result replicates the result of Corlett *et al.* (2004) because the mean difference in causal rating was in the same direction and the *p*-values were comparable (they found $p = .08$ using a one tailed *t* test, which would correspond to $p = .16$ had they used a two tailed *t* test as per our analysis). Nevertheless, although we replicate their empirical result, we argue against their *interpretation* of their result – a second failure to reach statistical significance suggests that their test statistic was not showing "a trend for a difference" (Corlett et al., 2004, p. 880), rather the most plausible interpretation of two failures to reach statistical significance is that the studies do not provide evidence for an overall effect of retrospective revaluation.

Our second replication followed Corlett *et al.* (2007) in testing for unovershadowing by comparing the Test 3 causal rating for cue D and cue J. The mean causal rating for cue D (M = 43.26, SD = 39.15) was higher than for cue J (M = 5.37, SD = 12.38), and this difference was statistically significant [$t(37) = 5.696$, $p < .001$]. This is difference is consistent with unovershadowing having occurred (because cue D went through unovershadowing training, but cue J did not). Empirically, this result replicates the result of Corlett *et al.* (2007) because the mean difference is in the same direction and the *p* values are comparable (they also found $p < .001$ using a two tailed *t* test). Nevertheless, although we replicate their empirical result, we disagree with their *interpretation* of their result – as discussed in our introduction, their comparison between cue D and cue J conflates unovershadowing and simple associative learning. As such these results should not be interpreted as evidence of unovershadowing, particularly in light of the results of the comparisons we report next.

Third, and most crucially, we tested for backward blocking and unovershadowing independently of each other using comparisons that control for simple associative learning. Unlike Corlett and colleagues' allergist task, our allergist task included measures of the causal rating of cues prior to retrospective revaluation training during Stage 2, which enabled us to test directly for a change in causal ratings in cue B (the cue that went through backward blocking training) and cue D (the cue that went through unovershadowing training) while controlling for simple associative learning during Stage 1 training. We tested for backward blocking by comparing the causal ratings for cue B at Test 2 and cue B at Test 3. The mean causal rating for cue B was higher at Test 2 (M = 59.32, SD = 30.77) than at Test 3 (M = 35.57, SD = 34.47), and this difference was statistically significant [$t(37) = 3.892$, $p < .001$]. This result is consistent with backward blocking having occurred. We tested for unovershadowing by comparing the causal rating for cue D at Test 2 against cue D at Test 3. The mean causal rating for cue D at Test 2 (M = 50.53, SD = 34.98) was higher at than at

Test 3 (M = 43.26, SD = 39.15), and this result was not statistically significant [$t(37) = 1.374$, $p = .178$]. This result does not provide support for unovershadowing having occurred – in fact, the mean rating for cue D numerically *decreased* after unovershadowing training, which is in the opposite direction to what we would expect if unovershadowing had occurred.

4.5 Discussion

In the present study we used two tests for an overall retrospective revaluation effect. We repeated the post-training simple comparison method used by Corlett *et al.* (2004). That is, we compared participants' final (Test 3) causal ratings for cues B and D. We found that, on average, cue D was rated higher than cue B, but this apparent difference was not statistically significant, which replicates the analogous result of Corlett *et al.* (2004). And, in addition to replicating their post-training comparison, we measured participants' causal ratings for B and D prior to Stage 2 (in Test 2) and after Stage 2 (in Test 3), which we used to test for a more sensitive interaction contrast to examine evidence for an overall retrospective revaluation effect. This analysis tested whether the difference in participants' ratings for cues B and D changed across Stage 2 training (i.e. between Test 2 and Test 3), which yielded a statistically significant interaction, indicating an overall retrospective revaluation effect, and replicating several earlier observations of this effect (e.g. Larkin et al, 1998).

Crucially, our pre/post experimental design also allowed us to investigate whether the overall retrospective revaluation effect was a consequence of backward blocking, or unovershadowing, or both – something that the experimental design used by Corlett *et al.* (2004) could not do. We did this by separately investigating the effect of Stage 2 training on cues B and D. We found that the mean causal ratings of each of these cues decreased after Stage 2. This decrease was statistically significant for cue B, but not cue D, which provided evidence that backward blocking occurred but did not provide evidence that unovershadowing occurred (in fact, the mean rating for cue D was *lower* after Stage 2 training, which is in the opposite direction to that predicted if unovershadowing had occurred). That is, these results suggest that the overall effect of retrospective revaluation suggested by the interaction contrast was driven by backwards blocking, not unovershadowing. This pattern of results does not support Corlett and colleagues' claim that the putative "trend for a difference" from their first study was caused primarily by unovershadowing.

In contrast to Corlett *et al.* (2004), which only reported a test for an overall retrospective revaluation effect, Corlett *et al.* (2006, 2007) only reported tests for an independent effect of unovershadowing. In none of these three studies was a test for an independent effect of backward blocking reported. Unfortunately, as we have argued in our introduction, existing literature does not support their justification for focusing exclusively on unovershadowing. Moreover, our analyses cast doubt on Corlett *et al.*'s (2007) empirical evidence for unovershadowing. They found that the causal rating for cue D at Test 3 was rated significantly higher than cue J at Test 3, and interpreted this as evidence for unovershadowing. However, as we have argued in our introduction, this comparison confounds simple associative learning with unovershadowing. The importance of this conflation was demonstrated in our study: just like Corlett *et al.* (2007), we found a highly statistically significant result when we compared cues D and J at Test 3 (which does not control for simple associative learning); but we did not find a statistically significant evidence when we compared cue D at Test 3 and Cue D at Test 2 (which does control for simple associative learning). In response to the critique by Griffiths *et al.* (2014), Corlett and Fletcher (2015) conceded that "Griffiths *et al.* do highlight an important inadequacy in our choice of controls" (p. 101), but suggest that their interpretations of the data are nevertheless valid. In response, Griffiths, Le Pelley, & Langdon (2015) emphasised that Corlett *et al.* (2006, 2007) "did not analyse their neutral data in a way that adequately measures a retrospective revaluation effect" (p. 107) and that these studies "had a contrast built into them... [that] controls elegantly for all factors other than belief revision" (p. 107). To date, these data have not been reanalysed using these more appropriate contrasts. As such, the results of the current study make a crucial contribution to this debate because, in a similar study with a much larger sample size, no evidence for retrospective revaluation was found. Given that our results provide no evidence for unovershadowing we suggest that there is a need for further allergist task studies that vary parameters to investigate the circumstances under which unovershadowing might be robustly observed. Only when such research has been undertaken will it be appropriate to start collecting data from clinical populations to investigate whether participants with clinical delusions show different patterns of unovershadowing to healthy controls.

4.6 Conclusion

We agree with Corlett and colleagues that adapting the associative learning allergist task for use in an fMRI brain scanner could be of considerable use for studying delusions in a manner that integrates cognitive and neurobiological levels of analysis. We also agree that abnormal prediction error processing may well play an important role in the formation and maintenance of delusions (see Coltheart et al., 2011; Miyazono, Bortolotti, & Broome, 2015). Nevertheless, our reassessment of the arguments and statistical comparisons that they presented in their studies, and our analyses of our new study that used a closely matched behavioural task, have revealed serious concerns about the extent to which their studies provide support for the prediction error theory of delusions. In particular, we find that their evidence that unovershadowing occurred in their three experiments is highly particularly problematic. Because their interpretation of their neuroimaging data relies crucially on the claim that unovershadowing occurred in their experiments, it is not clear whether changes in neural activity reported in the studies index prediction error signalling, or something else. This is troubling because these neuroimaging studies are routinely cited as providing the strongest evidence linking prediction error abnormalities to delusions (e.g., Clark, 2013;

Coltheart et al., 2011; Corlett, Frith, et al., 2009; Corlett et al., 2011, 2010; Corlett, Krystal, et al., 2009; Fletcher & Frith, 2009; Hohwy, 2013). While we agree that neuroimaging studies of prediction error would be well served by utilising existing associative learning procedures, we argue that it is important that such studies carefully consider and rule out simple associative explanations before attributing effects to complex "higher order" associative interactions.

4.6 References

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181-204. doi: 10.1017/S0140525X12000477

Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional belief. *Annual Review of Psychology*, *62*(5), 271–298. doi:10.1146/annurev.psych.121208.131622

Corlett, P. R., Aitken, M. R. F., Dickinson, A., Shanks, D. R., Honey, G. D., Honey, R. A. E., … Fletcher, P. C. (2004). Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron*, *44*(5), 877–888. doi:10.1016/j.neuron.2004.11.022

Corlett, P. R., Frith, C. D., & Fletcher, P. C. (2009). From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology*, *206*(4), 515–530. doi:10.1007/s00213-009-1561-0

Corlett, P. R., & Fletcher, P. C. (2015). Delusions and prediction error: clarifying the orles of behavioural and brain responses. *Cognitive Neuropsychiatry, 20*(2), 95-105 . doi:10.1080/13546805.2014.990625

Corlett, P. R., Honey, G. D., Aitken, M. R. F., Dickinson, A., Shanks, D. R., Absalom, A. R., … Fletcher, P. C. (2006). Frontal responses during learning predict vulnerability to the psychotogenic effects of ketamine: linking cognition, brain activity, and psychosis. *Archives of General Psychiatry*, *63*(6), 611. doi:10.1001/archpsyc.63.6.611

Corlett, P. R., Honey, G. D., Krystal, J. H., & Fletcher, P. C. (2011). Glutamatergic models of psychoses: prediction error, learning, and inference. *Neuropsychopharmacology*, *36*, 294–315. doi:10.1038/npp.2010.163

Corlett, P. R., Krystal, J. H., Taylor, J. R., & Fletcher, P. C. (2009). Why do delusions persist? *Frontiers in Human Neuroscience*, *3*. doi:10.3389/neuro.09.012.2009

Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R. F., Shanks, D. R., Robbins, T. W., … Fletcher, P. C. (2007). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain*, *130*, 2387–2400. doi:10.1093/brain/awm173

Corlett, P. R., Simons, J. S., Pigott, J. S., Gardner, J. M., Murray, G. K., Krystal, J. H., & Fletcher, P. C. (2009). Illusions and delusions: relating experimentally-induced false memories to anomalous experiences and ideas. *Frontiers in Behavioral Neuroscience*, *3*(53), 1–9. doi:10.3389/neuro.08.053.2009

Corlett, P. R., Taylor, J. R., Wang, X. J., Fletcher, P. C., & Krystal, J. H. (2010). Toward a neurobiology of delusions. *Progress in Neurobiology*, *92*(3), 345–369. doi:10.1016/j.pneurobio.2010.06.007

David, A. S. (1999). On the impossibility of defining delusions. *Philosophy, Psychiatry, & Psychology*, *6*(1), 17–20.

De Houwer, J., & Beckers, T. (2002). Second-order backward blocking and unovershadowing in human causal learning. *Experimental Psychology*, *49*(1), 27–33. doi:10.1027/1618-3169.49.1.27

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. *The Quarterly Journal of Experimental Psychology: Section B*, *49*(1), 60–80.

Fletcher, P. C., Anderson, J. M., Shanks, D. R., Honey, R., Carpenter, T. A., Donovan, T., … Bullmore, E. T. (2001). Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nature Neuroscience*, *4*, 1043–1048. doi:10.1038/nn733

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*(1), 48–58. doi:10.1038/nrn2536

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, *159*(3), 417–458. doi:10.1007/s11229-007-9237-y

Graham, G. (2010). *The disordered mind: an introduction to philosophy of mind and mental illness*. New York, NY: Routledge.

Griffiths, O., Langdon, R., Le Pelley, M. E., & Coltheart, M. (2014). Delusions and prediction error: re-examining the behavioural evidence for disrupted error signalling in delusion formation. *Cognitive Neuropsychiatry, 19*(5), 439-467. doi:10.1080/13546805.2014.897601

Griffiths, O., Le Pelley, M. E., & Langdon, R. (2015). The bridge between neuroscience and cognitive must be tethered at both ends. *Cognitive Neuropsychiatry, 20*(2), 106-108. doi:10.1080/13546805.2014.993464

Hohwy, J. (2013). *The predictive mind*. Oxford, UK: Oxford University Press.

Howes, O. D., & Kapur, S. (2009). The dopamine hypothesis of schizophrenia: version III—the final common pathway. *Schizophrenia Bulletin*, *35*(3), 549–562. doi:10.1093/schbul/sbp006

Larkin, M. J. W., Aitken, M. R. F., & Dickinson, A. (1998). Retrospective revaluation of causal judgments under positive and negative contingencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1331–1352.

Le Pelley, M. E., & McLaren, I. (2001). Retrospective revaluation in humans: Learning or memory? *The Quarterly Journal of Experimental Psychology: Section B*, *54*(4), 311–352. doi:10.1080/02724990143000072

Mather, M., Cacioppo, J. T., & Kanwisher, N. (2013). Introduction to the special section: 20 years of fMRI--What has it done for understanding cognition? *Perspectives on Psychological Science*, *8*(1), 41–43. doi:10.1177/1745691612469036

Miyazono, K., Bortolotti, L., & Broome, M. R. (2015). Prediction-error and two-factor theories of delusion formation: competitors or allies? In N Galbraith (Ed.), *Aberrant Beliefs and Reasoning* (pp. 34-54). East Sussex, UK Psychology Press.

Shallice, T., & Cooper, R. P. (2011). *The organisation of mind*. Oxford: Oxford University Press.

Shanks, D. R. (2010). Learning: From association to cognition. *Annual Review of Psychology*, *61*, 273–301. doi:10.1146/annurev.psych.093008.100519

Turner, D. C., Aitken, M. R. F., Shanks, D. R., Sahakian, B. J., Robbins, T. W., Schwarzbauer, C., & Fletcher, P. C. (2004). The role of the lateral frontal cortex in causal associative learning: exploring preventative and super-learning. *Cerebral Cortex*, *14*(8), 872–880. doi:10.1093/cercor/bhh046

Wasserman, E. A., & Berglan, L. R. (1998). Backward blocking and recovery from overshadowing in human causal judgment: the role of within-compound associations. *The Quarterly Journal of Experimental Psychology: Section B*, *51*(2), 121–138. doi:10.1080/027249998393357

Wasserman, E. A., Kao, S. F., Van Hamme, L. J., Katagiri, M., & Young, M. E. (1996). Causation and association. In *The psychology of learning and motivation: Vol 43* (pp. 207–264). San Diego, CA: Academic Press.

Chapter Five: Population structure and cultural geography of a folktale in Europe

5.1 Abstract

Despite a burgeoning science of cultural evolution, relatively little work has focused on the population structure of human cultural variation. By contrast, studies in human population genetics use a suite of tools to quantify and analyse spatial and temporal patterns of genetic variation within and between populations. Human genetic diversity can be explained largely as a result of migration and drift giving rise to gradual genetic clines, together with some discontinuities arising from geographical and cultural barriers to gene flow. Here, we adapt theory and methods from population genetics to quantify the influence of geography and ethnolinguistic boundaries on the distribution of 700 variants of a folktale in 31 European ethnolinguistic populations. We find that geographical distance and ethnolinguistic affiliation exert significant independent effects on folktale diversity and that variation between populations supports a clustering concordant with European geography. This pattern of geographical clines and clusters parallels the pattern of human genetic diversity in Europe, although the effects of geographical distance and ethnolinguistic boundaries are stronger for folktales than genes. Our findings highlight the importance of geography and population boundaries in models of human cultural variation and point to key similarities and differences between evolutionary processes operating on human genes and culture.

5.2 Introduction

Parallels between processes of genetic and cultural evolution (Atkinson & Gray, 2005; Mesoudi, Whiten, & Laland, 2004) mean that method and theory developed to analyse biological data can be used to study language, culture and the archaeological record (Bentley, Hahn, & Shennan, 2004; Eerkens & Lipo, 2007; Gray, Greenhill, & Ross, 2007; Mesoudi, Whiten, & Laland, 2006; O'Brien & Lyman, 2003). A major focus of empirical research on cultural variation and change has been the analysis of data coding for the presence or absence of population-level cultural traits across ethnolinguistic groups (groups defined along ethnic and/or linguistic lines). It has been argued that these traits are frequently transmitted with a high degree of fidelity down ethnolinguistic lineages, analogous to genetic inheritance in biological species, supporting what has been dubbed the "cultures as species" model (Pagel & Mace, 2004). Language change may be a paradigm example of such "species-like" cultural evolution (Pagel, 2009), and language family trees inferred using phylogenetic methods are now routinely used as lineages on which to model the evolution of a wide variety of population-level cultural traits (Currie, Greenhill, Gray, Hasegawa, & Mace, 2010; Fortunato, Holden, & Mace, 2006; Jordan, Gray, Greenhill, & Mace, 2009).

However, cultures do not always behave like species (Borgerhoff Mulder, Nunn, & Towner, 2006; Boyd, Richerson, Borgerhoff Mulder, & Durham, 1997; Moore, 1994). Characterising ethnolinguistic groups as having population-level cultural traits can be problematic when there is significant heterogeneity within groups. Furthermore, while horizontal transmission of genes between species is rare, the exchange of cultural traits between ethnolinguistic groups is not. As Boyd *et al.* (1997) argue, there exists a spectrum of possibilities for the degree of coherence of culture within ethnolinguistic groups, ranging from core cultural traditions with less cohesive peripheral aspects, to assemblages of bounded cultural packages lacking core traditions, to mere collections of ephemeral and unbounded cultural traits. Where within-population variation and horizontal transmission are high, a macro-evolutionary "cultures as species" model provides, at best, an incomplete picture that ignores internal diversity and the micro-evolutionary processes shaping patterns of variation. There is thus a need for research methods that quantify, rather than ignore, within-population and spatial variation in culture.

Population geneticists have developed a suite of tools for characterizing patterns and processes of genetic variation within species owing to mutation, selection, gene flow and drift (Kimura, 1983; Maynard Smith, 1998). Wright's *F* statistic (F_{ST}; Wright, 1943) and associated metrics such the *phi* statistic (Φ_{ST}; Excoffier, Smouse, & Quattro, 1992) are routinely used to measure how variance in genetic diversity is partitioned within and between populations. The F_{ST} quantifies the relative variation of traits within versus between populations and is calculated as the correlation of randomly chosen variants within a population relative to a similar correlation across the meta-population (Holsinger & Weir, 2009). An F_{ST} or Φ_{ST} value of 0 indicates no differentiation between populations, whereas a value of 1 indicates complete differentiation.

Analyses of autosomal single nucleotide polymorphisms (SNPs) in human populations around the world have yielded average F_{ST} estimates of between 0.052 and 0.130 (Barbujani & Colonna, 2011), indicating that, on a global scale, roughly 5–13% of human autosomal genetic variation occurs between populations. Between-population variation can be much lower when examining genetic diversity within continents, particularly in Europe (Cavalli-Sforza, Menozzi, & Piazza, 1994; Li et al., 2008). Recently, high-resolution studies of SNP data from European populations have found extremely low average F_{ST} estimates of 0.0025–0.004 between populations (Lao et al., 2008; Novembre et al., 2008). Low levels of genetic diversity between human populations have been used to argue against the validity of the biological concept of race (Lewontin, 1972; Marks, 2010) and against the feasibility of genetic group-level selection in humans (Bell, Richerson, & McElreath, 2009).

Another important line of inquiry in population genetics uses spatial analysis of human genetic diversity to shed light on the processes shaping our gene pool. Although human genetic variation falls into a number of regional clusters, the predominant pattern is clinal, with much of the apparent regional clustering attributable to discontinuous spatial sampling (Handley, Manica, Goudet, & Balloux, 2007). Genetic distance between human populations increases with geographical distance at both continental and global scales, and across a variety of markers (Cavalli-Sforza et al., 1994). A smooth clinal pattern of genetic variation is often taken to support an "isolation by distance" (IBD) model (Handley et al., 2007; Wright, 1943) in which individuals tend to migrate short distances between neighbouring populations, taking their genes with them, resulting in gradual diffusion of genetic variants across the landscape. In Europe, an IBD model is supported by a remarkable fit between genes and geography: a recent study of high- resolution autosomal SNP data found that the first and second principal components of genetic variation recreated a map of the continent, albeit explaining only a small percentage of the overall variation (0.30% and 0.15%, respectively) (Novembre et al., 2008). Conversely, departures from a clinal pattern of human genetic variation expected under the IBD "null" model have been used to identify population boundaries, prehistoric migrations and ancient selection pressures (Barbujani & Sokal, 1990; Itan, Powell, Beaumont, Burger, & Thomas, 2009; Novembre & Stephens, 2008).

Research on human population structure and spatial variation has allowed population geneticists to gain insights into human prehistory and the processes operating within populations that give rise to global patterns of genetic diversity (Handley et al., 2007; Li et al., 2008). It has long been argued by anthropologists and archaeologists that research on cultural evolution also needs to take "population thinking" seriously (Bentley et al., 2004; Boyd & Richerson, 1985; Neiman, 1995). By quantifying population structure and spatial variation in cultural diversity, we can learn how micro-scale processes operating within populations act to shape macro-scale between-population variation in human culture.

Recently, scholars have begun to borrow theory and analytical tools from population genetics to study cultural variation within populations. Random copying models analogous to Kimura's (1983) neutral genetic drift model have been used to predict variation and change in the archaeological record (Eerkens & Lipo, 2005; Kempe, Lycett, & Mesoudi, 2012; Lycett, 2008) and in contemporary culture (Bentley et al., 2004). Bell *et al.* (2009) used cross-cultural data from the *World Values Survey* to calculate pairwise cultural F_{ST} values for 150 neighbouring countries, which they compared with previously published genetic F_{ST} values (like the genetic F_{ST}, cultural F_{ST} is a measure of the relative variation of traits within versus between populations). They found that the average cultural F_{ST} value between neighbouring countries (mean = 0.080) was an order of magnitude larger than the average SNP genetic F_{ST} value between the same countries (mean = 0.0053), which they argued demonstrates a greater potential for group selection on culture than genes. Rzeszutek, Savage, and Brown (2011) examined cross-cultural variation in song characteristics across 16 Formosan-speaking ethnolinguistic groups and found an overall Φ_{ST} of 0.02, indicating that approximately 2% of variation was between populations. In addition, debates in experimental economics have begun to focus on within- versus between-population variation in strategies employed in economic games (Gachter, Herrmann, & Thoni, 2010; Henrich et al., 2012; Lamba & Mace, 2012).

While these studies make important first steps towards quantifying cultural variation within and between populations, none of them investigated how this variation is patterned spatially. This renders estimates of population structure (such as F_{ST} and Φ_{ST} values) difficult to interpret because, as research on human genetic diversity has repeatedly demonstrated, apparent population structure can be an artifact of discontinuous spatial sampling, rather than group boundaries (Handley et al., 2007). There is therefore a need for research that quantifies the independent effects of group boundaries and geography on patterns of cultural diversity and examines when and why these patterns vary across different elements of culture.

Here, we adapt tools from population genetics to quantify the influence of both population structure and geography on 700 variants of the folktale *The Tale of the Kind and the Unkind Girls* (Roberts, 1958), drawn from 31 ethnolinguistic populations across Europe shown in Figure 5.1. Described by the folklorist Thompson (1946) as "one of the most popular of oral tales" (p. 126), versions of this folktale are found all over Europe. Two variants appear in the Brothers Grimm fairy tale collection (*Die drei Männlein im Walde* and *Frau Holle*), and a motif was used by Shakespeare in *The Merchant of Venice*. Variants of the folktale typically tell a moralistic story of a kind girl who is rewarded for her generosity and an unkind girl who is punished for her selfishness.

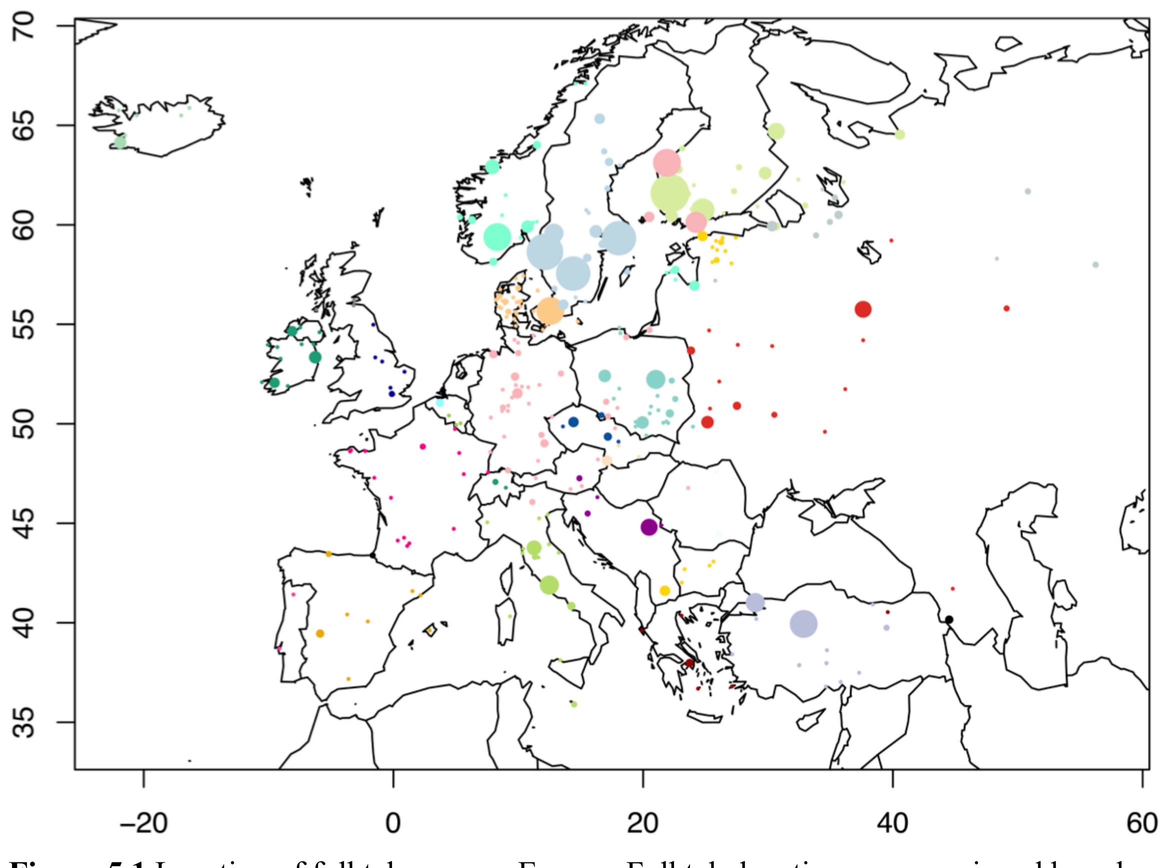


Figure 5.1 Location of folktales across Europe. Folktale locations were assigned based on information provided by Roberts (1958). Points are coloured to show different ethnolinguistic groups. The size of each point is proportional to the number of variants sampled from that location.

The folktale has been summarised as follows:

> [T]he story is concerned with two girls, one of whom is good and kind while the other is evil and unkind. The good girl leaves her home and sets out on a journey for some reason. Her bucket may fall into a well and she climbs after it, or she may pursue a rolling cake. During her journey, in one important form of the story, she usually meets a cow, an apple tree, and an oven which ask her to help them. She complies with these requests and continues her journey until she comes to a house. These encounters on the way are absent from the second important form of the tale. At the house she takes services with an old woman or witch and performs housekeeping chores and other tasks. At the end of a year the girl wishes to return to her home. As a recompense for her labor the old woman offers the girl her choice between several boxes. The girl modestly chooses the smallest and least attractive box. When she reaches home and opens her box she finds it is full of gold. The bad girl is jealous and resolves to try her fortune. She sets out in the same way that the heroine did. She haughtily refuses to help the animals or things she meets on the way and at the house either refuses to work altogether or does a very poor job. She greedily chooses the biggest box, which when opened at home, is found to be full of snakes. It must be understood that the above outline is only a rough generalization and that there are innumerable variations upon this simple theme. (Roberts, 1958, p. 3)

There are a number of features of this folktale dataset that make it particularly attractive for studying the influence of population structure and geography on cultural variation. First, the dataset includes multiple samples of folktale variants drawn from the same ethnolinguistic group, allowing the quantification of within- versus between-group variation. Second, the dataset includes geographical information for 84% of the folktale variants, which affords an opportunity to disentangle effects of group membership and geography. Third, most of the folktale variants included in the dataset were collected during the late nineteenth and early twentieth centuries, before communication technology and air travel transformed how ideas and people spread. Fourth, given that variation in this folktale was likely to have been predominantly selectively neutral (i.e. not "functional" in the sense of being tested against the natural environment; Boyd & Richerson, 1985; Neiman, 1995), it may provide a plausible "null" model of cultural diffusion, akin to IBD in population genetics, against which the effects of selection, population boundaries and cultural ancestry can be tested. Finally, the folktale variants in the dataset were independently coded for narrative content by a noted folklore scholar (Roberts, 1958) according to the well-established historic–geographic method of folklore analysis (Thompson, 1946).

We examine the independent effects of population structure and geography on variation in this folktale across Europe using three stages of analysis. First, we quantify individual folktale variation within versus between ethnolinguistic groups and examine whether between-population folktale variation is greater than between-population genetic variation, as has been found for other cultural traits (Bell et al., 2009). Second, we investigate the processes underlying any between population differences. We test whether individual folktale variation shows a predominantly clinal pattern, like that observed in human population genetic variation (Handley et al., 2007), and quantify the independent effects of geography and ethnolinguistic affiliation. Third, we examine how the various folktale populations cluster in Europe, using pairwise population Φ_{ST} distances. We ask whether these populations show a hierarchical, tree-like pattern of branching, probably reflecting sequential colonisation and vertical inheritance of coherent (perhaps linguistic) lineages, or a more reticulate pattern, aligned to geography, suggesting a process of local diffusion.

5.3 Material and methods

5.3.1 Data

We sourced folktale data from Roberts' (1958) study of *The Tale of the Kind and the Unkind*

Girls—tale type 480 according to the *Aarne–Thompson–Uther Tale Type Index* (Uther, 2004). Roberts indicated the presence and absence of important narrative elements in each folktale variant using multistate character codings according to principles of the *Historic–Geographic Method* of folklore analysis (Thompson, 1946). There were clearly some minor typographic errors and inconsistency in the data, and the data were not presented clearly as tables. When we encountered typographic errors and inconsistencies we attempted to infer correct values. We summarise the corrected data as counts for each level of the characters in the appendix. For example, one coded narrative element is the location where the main protagonist meets some other key characters, with the location coded according to 12 character-states, including at the bottom of a well, by a river, in a field, on a mountain-side and in a cave.

We assigned folktale variants to populations using the ethnolinguistic assignments provided by the source dataset. We analysed only those folktale variants that were drawn from ethnolinguistic populations in Europe because many of the other geographical regions were poorly sampled and included folktale variants that might reflect more recent post-colonial movements rather than long-standing geographical and ethnolinguistic patterns (Roberts, 1958). In total, our analysis included 700 folktale variants drawn from 31 European ethnolinguistic populations, with a mean of 23 folktale variants per population (Armenian, 3; Basque, 2; Bulgarian, 8; Czech, 11; Danish, 48; English, 8; Estonian, 16; Finnish, 83; Swedish in Finland, 25; Flemish, 6; French, 16; German, 61; Greek, 11; Icelandic, 11; Irish, 22; Italian, 33; Latvian, 13; Norwegian, 48; Polish, 45; Portuguese, 2; Romanian, 4; Russian, 32; Finno-Ugric in Russia, 23; Scottish, 3; Slovenian, 6; Spanish, 11; Swedish, 101; Swiss German, 3; Turkish, 32; Walloon, 3; Yugoslavian, 13).

We recoded the presence or absence of narrative elements as “1” or “0”, respectively, to produce a matrix of 700 folktale variants coded across 393 binary traits (traits coded as “other” were excluded because it is a catchall category such that a shared presence of “other” does not represent similarity). For analysis, this presence/absence matrix was converted to a Jaccard distance matrix reflecting pairwise distances between all folktale variants. The Jaccard distance for each pair of folktale variants was calculated as the sum of the number of traits that are present in one variant but not the other, divided by the sum of the number of traits that are present in one or both of the variants. The Jaccard distance is particularly appropriate for analysing this cultural dataset because it standardises for the number of traits observed for each pair and shared absences do not contribute to similarity (Rogers & Ehrlich, 2008).

The geographical locations of the folktale variants are shown in Figure 4.1. They were estimated using locality information included in the source dataset. 16% of the folktale variants did not include locality information beyond ethnolinguistic affiliation. For these folktale variants, geographical coordinates were assigned as the centroid location of the points sampled from the ethnolinguistic group to which they belonged. Removing these cases from the analysis did not qualitatively affect any of the results we report. Geographical coordinates were used to calculate pairwise geographical distance and logged geographical distance matrices between individual folktale variants, and between the 31 ethnolinguistic populations (using the centroid of geographical coordinates for each population). Pairwise distances were calculated using great circle distances in GenAlEx version 6.4 (Peakall & Smouse, 2006).

We used linguistic divergence between ethnolinguistic groups to index cultural ancestry. A language dissimilarity matrix was calculated using patristic distances between Indo-European languages inferred from Gray and Atkinson's (2003) phylogenetic analysis of the Indo-European language family. All Indo-European ethnolinguistic populations included in the folktale dataset were represented in Gray and Atkinson's analysis, with the exception of Scottish. Nevertheless, Scottish can be reliably placed as a close sister language to Irish in the Indo-European tree (Gillies, 1993) so we assigned Scottish the same distance as Irish to all languages (except to Irish itself, which was assigned a distance equivalent to the minimum distance between languages observed in the initial data). Assigning distances to languages outside the Indo-European family is more problematic. Higher-level language family groupings have been proposed, but they remain highly controversial (Campbell & Posner, 2008), making precise estimates of distances between languages from different language families unfeasible. To generate approximate values, we set distances between languages from different language families (Indo-European, Turkic and Finno-Ugric) to 1.25 times the maximum observed distance between Indo-European languages. The ethnolinguistic category "Finno-Ugric in Russia" was also problematic, because the particular Finno-Ugric languages were not recorded. Because Finno-Ugric shows a comparable level of internal diversity to Indo-European (Marcantonio, 2002), we set a distance for languages within the Finno-Ugric language family (Finnish, Estonian and "Finno-Ugric speakers in Russia") to the average distance between languages in the Indo-European language family. We found our results were robust across a range of between-family distance multipliers from 1 to 3.

An ethnolinguistic identity matrix for individual folktale variants was created by scoring the distance between folktale variants as "0" if they came from the same ethnolinguistic group and "1" if they came from a different group. These usually correspond to language speaker populations (e.g. Spanish), but twice to subpopulations within a language (Swiss German, Swedish speakers in Finland) and once to a group of related languages (Finno-Ugric speakers in Russia).

5.3.2 Analysis

Cultural population structure across ethnolinguistic groups was investigated using the analysis of molecular variance (AMOVA) (Excoffier et al., 1992) technique as implemented in Arlequin version 3.5.1.2 (Excoffier, Laval, & Schneider, 2005). AMOVA provides a measure of the proportion of variance within versus between populations using between-population Φ_{ST} values—a value of 0 indicates no differentiation between populations, whereas a value of 1 indicates complete differentiation. Unlike the F_{ST} statistic, which is based on variant frequencies, the Φ_{ST} statistic extracts additional information from the data by accounting for distances between variants. The method takes as input a pairwise matrix of distances between sampled variants, together with information on the population each variant was sampled from. Because AMOVA makes no assumptions about the units of analysis or the mechanisms generating diversity, it is equally suited to analysing cultural data from ethnolinguistic groups or genetic data from biological populations. Although geneticists use a measure of genetic distances between sequences, here we use our Jaccard distance matrix of distances between folktale variants. By calculating pairwise population Φ_{ST} values across ethnolinguistic groups, it is possible to quantify the average level of within- versus between-group variation, as well as population pairwise distances. Negative Φ_{ST} values have no interpretation and, following standard practice, were set to zero. Statistical significance of Φ_{ST} values was tested using 1000 random permutations.

Spatial autocorrelations among (i) individual folktale variants and (ii) pairwise Φ_{ST} values for ethnolinguistic populations were calculated using the method implemented in GenAlex version 6.4 (Peakall & Smouse, 2006). This autocorrelation method uses a pairwise geographical distance matrix and a pairwise folktale distance matrix to calculate an autocorrelation coefficient *r* across a specified range of geographical distance classes. The autocorrelation coefficient provides a measure of the similarity between pairs of folktales whose geographical separation falls within each distance class. Tests for statistical significance were performed using two methods, calculating *r* across 1000 random permutations and 1000 bootstrap estimates (Peakall, Ruibal, & Lindenmayer, 2003).

In order to investigate the independent effects of geography, ethnolinguistic affiliation and cultural ancestry on variation in individual folktale variants, we calculated correlations and partial correlations between the folktale, geographical, linguistic and ethnolinguistic identity distance matrices using Mantel and partial Mantel tests (Mantel, 1967; Smouse, Long, & Sokal, 1986) in Arelquin version 3.5.1.2 (Excoffier et al., 2005), with significance assessed using 1000 random permutations. We used the same approach to test for correlations between geographical distance, linguistic distance and pairwise Φ_{ST} values between ethnolinguistic populations.

In order to visualise the pattern of relationships between populations and identify population clusters, we constructed a NeighbourNet (Bryant & Moulton, 2004) in SplitsTree version 4.11.3 from the folktale pairwise population Φ_{ST} values. The NeighbourNet algorithm is useful for identifying complex transmission histories of population divergence and convergence (Bryant, Filimon, & Gray, 2005; Gray, Bryant, & Greenhill, 2010). The method does not assume a simple tree-like model of evolution; instead, evidence for such a model appears as bifurcating "tree-like" splits in the graph. Conversely, evidence for convergence or horizontal transmission owing to cultural borrowing will appear as reticulate, "box-like" structures representing conflicting population subdivisions.

5.4 Results and discussion

5.4.1 Population structure

Our AMOVA reveals moderate but highly significant population structure in folktale variation across the sampled ethnolinguistic groups, with 9.1% of the variation among individual folktales occurring between populations (average $\Phi_{ST} = 0.091$, $p < .001$). Some of the ethnolinguistic groups in our dataset had small sample sizes, which can result in unreliable Φ_{ST} values. To investigate whether they may have biased our results, we repeated the AMOVA with small populations (less than five variants) removed. Consistent with the full analysis, we again found 9% between-population variation ($\Phi_{ST} = 0.090$, $p < .001$). This value is comparable to levels of variation observed in attitudes and values between neighbouring nations (8%) (Bell et al., 2009) and to between-population behavioural variation in economic games (4–38%) (Gachter et al., 2010; Henrich et al., 2012).

A value of 9.1% is also within the range of between-population variation in global human autosomal genetic diversity, which range from 5% to 13% (Barbujani & Sokal, 1990). However, estimates of between-population genetic variation in comparable European populations range from 0.25% to 0.40% (Lao et al., 2008; Novembre et al., 2008). This order of magnitude difference in Europe fits with the finding that cultural F_{ST} scores calculated using variation in attitudes and values between neighbouring nations (F_{ST} = 0.08 or 8%) are higher than genetic F_{ST} scores for the same populations (F_{ST} = 0.005 or 0.5%) (Bell et al., 2009).

When comparing our results with estimates of human genetic diversity, it is important to note that, while each sampled genotype can be tied to an individual person, here we are not tracking characteristics (behavioural or genetic) of individual people—that is, we do not have information about which individuals in a population know which folktale variant(s). Although tracking the characteristics of individual people is appropriate for some cultural traits (Bell et al., 2009), it makes little sense for traits such as folktales because, unlike genes, one person can know many folktales and folktales can move without people. Instead, our approach tracks the cultural entities themselves, in effect treating individual folktale variants in ethnolinguistic groups like population geneticists treat genetically distinct haploid organisms in biological populations. Rzeszutek *et al.* (2011) used a similar approach in their analysis of Formosan song variants, although, interestingly, our estimate of between-population variation is closer to Bell *et al.*'s 8% than Rzeszutek *et al.*'s 2% (see section d for a possible explanation for this).

While AMOVA allows us to quantify variation between ethnolinguistic groups, it does not tell us whether the differences we observe are the result of measurable ethnolinguistic boundaries and divergence along cultural lineages, or purely clinal patterns of geographical variation, or some combination of the two. In order to determine how the between-population differences we observe arose, we first consider the effects of geography and then test for departures from a purely clinal model based on ethnolinguistic affiliation.

5.4.2 Geographical clines

Mantel tests on individual folktale data show clear clinal patterning (Table 5.1). Logged geographical distance is the best single predictor of folktale similarity, explaining 8.9 % of the variance ($r^2 = .089$, $p < .001$). Unlogged geographical distance explains 6.4 % of the variance ($r^2 = .064$, $p < .001$). By comparison, ethnolinguistic identity and language distance explain 6.8 % ($r^2 = .0683$, $p < .001$) and 2.6 % ($r^2 = .0262$, $p < .001$) of the variance in folktale similarity, respectively. Spatial autocorrelation analysis also shows a highly significant relationship between individual folktale distance and geographical distance (Figure 5.2a). Although the correlation is small, it is roughly an order of magnitude greater than observed in similar analyses of autosomal genetic distances between individuals across Europe (Lao et al., 2008; Novembre et al., 2008).

Table 5.1 Results of Mantel and partial Mantel tests of correlations between individual folktale Jaccard distance values, geographical distance, logged geographical distance, linguistic distance and ethnolinguistic group.

	Variance explained	*p* value
Individual folktale distance values (*n*=700) predicted by		
Geography	6.38%	<.001
Log(Geography)	8.92%	<.001
Ethnolinguistic group	6.83%	<.001
Language	2.62%	<.001
Log(Geography) controlling for Ethnolinguistic group	6.64%	<.001
Log(Geography) controlling for Language	8.50%	<.001
Ethnolinguistic group controlling for Log(Geography)	3.68%	<.001
Ethnolinguistic group controlling for Language	6.14%	<.001
Language controlling for Log(Geography)	0.54%	.006
Language controlling for Ethnolinguistic group	0.99%	<.001

Our population-level analyses also show clear clinal spatial structure (Table 5.2). Unlike the individual folktale analyses, geographical distance explains more of the variance in pairwise population Φ_{ST} values (14.8%, $r^2 = .148$, $p < .001$) than does logged geographical distance (13.0%, $r^2 = .130$, $p < .001$). By comparison, language distance explains 7.5% of the variance (r^2 = .0751, p = .014). These findings hold when populations with small sample sizes are excluded (Table 5.2). The shape and magnitude of spatial autocorrelation at the population level (Figure 5.2b) is similar to that found in analyses of human genetic variation between populations in Europe (Rosser et al., 2000).

Partial Mantel tests provide insights into the processes driving these spatial patterns of folktale variation at the individual level (Table 5.1). Logged geographical distance remains a significant predictor of individual folktale variation, even after controlling for ethnolinguistic identity (r^2 = .085, p < .001) and language distance (r^2 = .066, p < .001), explaining 8.5 % and 6.6 % of the variance respectively. This indicates that spatial patterning is not simply the result of cultural divisions (as measured by ethnolinguistic affiliation) or cultural ancestry (as measured by language distance). In fact, the strongest individual folktale correlations occur at distances of less than 200 km, suggesting highly localised within-group effects of geography on folktale variation (Figure 2a).

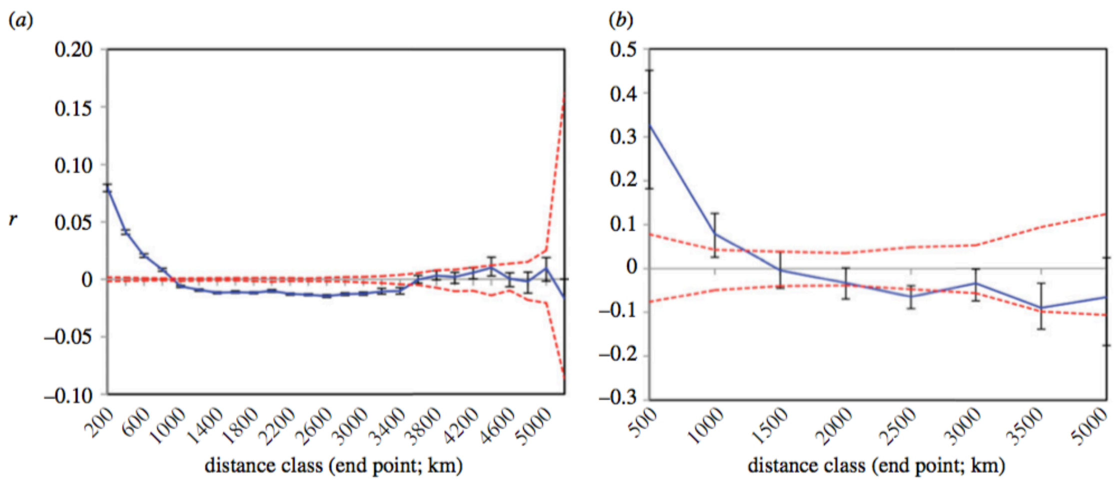


Figure 5.2 Folktale spatial autocorrelation analysis (Peakall & Smouse, 2006). Spatial correlogram plot showing correlation coefficient (r) as a function of distance for (a) individual-level data from 700 folktales using pairwise Jaccard distances and (b) population-level data from 31 ethnolinguistic groups using pairwise F_{ST} values. The permuted 95% CI (dashed lines) and the bootstrapped 95% confidence error bars are also shown.

Table 5.2 Results of Mantel and partial Mantel tests of correlations between population pairwise matrices of folktale F_{ST} values, geographical distance, logged geographical distance and linguistic distance.

	Variance explained	*p* value
Population folktale Φ_{ST} values (*n*=31) predicted by		
Geography	14.75%	<.001
Log(Geography)	12.99%	<.001
Language	7.51%	.014
Geography controlling for Language	12.61%	.005
Language controlling for Geography	4.33%	.099
Population folktale Φ_{ST} values [minus groups with <5 folktales] (*n*=23) predicted by		
Geography	16.78%	.001
Language	8.23%	.035
Geography controlling for Language	14.40%	.013
Language controlling for Geography	4.40%	.140

The importance of geography is reinforced at the population level (Table 5.2). Geographical distance explains 12.6 % of the variance in between-population F_{ST} values when controlling for language distance ($r^2 = .126$, $p < .001$), but language distance is not a significant predictor of Φ_{ST} values when controlling for geographical distance ($r^2 = .043$, $p < .106$). There are two plausible explanations for these results. First, the folktale and language histories are decoupled, either because the folktales spread much later than the spread of languages across Europe, or because any legacy of deep cultural ancestry inherited down language lineages has been obscured by subsequent folktale evolution and geographical diffusion. Second, both language distance and geography are linked to folktale evolution, but these predictors share common variance.

The NeighbourNet constructed from pairwise Φ_{ST} values between ethnolinguistic groups reveals a highly reticulate network and regional clustering of populations (Figure 5.3). This does not support the idea that current folktale variation is the result of a sequential colonisation of the landscape by vertically inherited, coherent cultural lineages (linguistic or otherwise). Convergent evolution of traits and/or trait reversals could account for some reticulation in the graph, but they would not be expected to generate the regional clustering we observe. Together, then, our individual and population-level results point to the primacy of local cultural diffusion processes between neighbouring folktale variants.

5.4.3 Ethnolinguistic boundaries

Measureable differences between groups do not necessarily point towards population structure since they could be the result of clinal variation that is masked by discontinuities in spatial sampling (Handley et al., 2007). On Boyd *et al.*'s (1997) spectrum of cultural descent types, this would suggest that folktales are simply diffusing across the landscape and are not part of coherent cultural traditions. By testing for departures from a purely clinal model, we can determine whether ethnolinguistic boundaries act as a barrier to the spread of folktales.

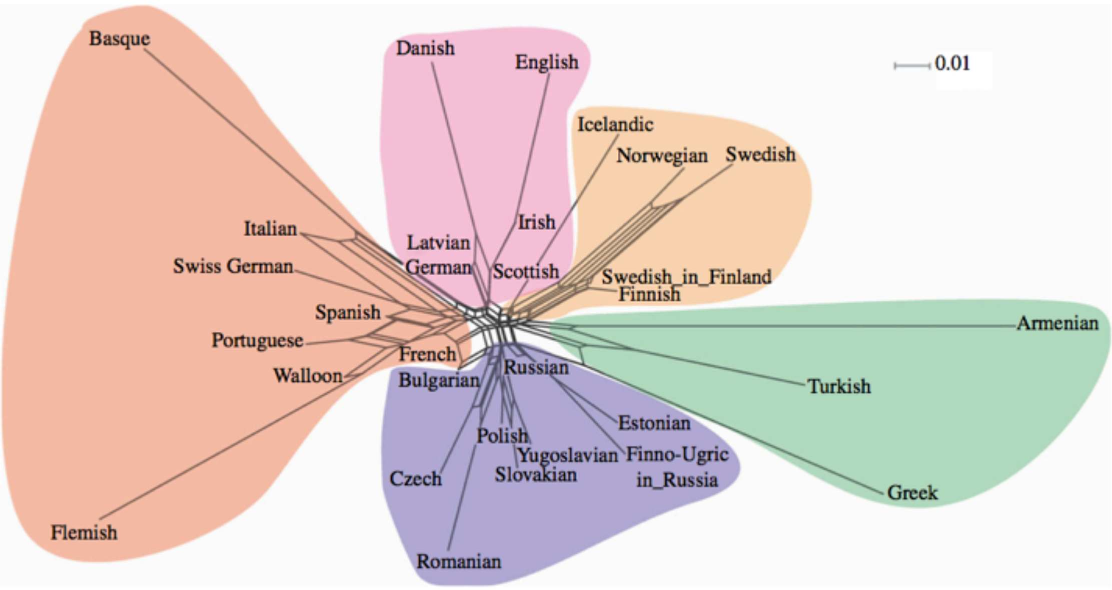


Figure 5.3 NeighbourNet (Bryant et al., 2005) of European folktale populations. The relationship between folktale populations across Europe, based on population folktale F_{ST} values. Populations that are closer together tend to have more similar folktales. Box-like structures show the reticulate nature of folktale similarity, indicating extensive horizontal transmission (as opposed to vertical transmission down cultural lineages). Shaded polygons show the five clusters discussed in the main text.

A partial Mantel test that uses ethnolinguistic identity to predict folktale variation while controlling for geography shows that ethnolinguistic identity explains a significant proportion of the variation in individual folktales, even after controlling for geographical distance ($r^2 = .037$, $p < .001$). Ethnolinguistic identity therefore represents a barrier to folktale transmission. Based on the regression coefficients from our model incorporating geographical distance and ethnolinguistic identity, we can infer that the magnitude of this cultural barrier effect is equivalent to multiplying geographical distance between folktale variants by a factor of 10 (the relationship is multiplicative, rather than additive, because we are using logged geographical distance). In other words, folktales from the same culture found 100 km apart are, on average, as similar as folktales found 10 km apart in different cultures.

Studies of human genetic diversity have likewise identified barriers to gene flow that may be related to ethnolinguistic identity (Barbujani & Colonna, 2010; Barbujani & Sokal, 1990). In both the folktale and genetic case, barriers could arise if there is a reduced probability of transmission across ethnolinguistic boundaries. If folktales cross ethnolinguistic boundaries less easily than genes, this could partly explain higher folktale F_{ST} values. However, in the case of folktales, another possibility is that cultural transmission biases operating within, but not across, ethnolinguistic groups may differentially impact which folktale elements are successfully copied. Content-dependent biases, such as favouring certain motifs for their meaning in certain cultures, or context-dependent biases, such as conformist or prestige bias (Boyd & Richerson, 1985; Henrich & Gil-White, 2001), could lead to highly successful variants that are particular to each group.

5.4.4 Patterns and processes of human cultural evolution

Our findings highlight key similarities and differences between patterns and processes of folktale and genetic variation in Europe. Like genetic variation, most folktale variation occurs within ethnolinguistic groups. However, across Europe, the folktales in our study show an order of magnitude more between-population variation than genes. Three factors are likely to be at work here. First, faster rates of cultural evolution could increase the likelihood of between-population differences arising (Gray et al., 2007)—although this also increases within-population variation. Second, the ethnolinguistic barrier effect we identify suggests that content- and/or context-dependent cultural transmission biases (Boyd & Richerson, 1985) are acting to limit information flow across group borders, suppress internal variation and/or accentuate group differences. Third, the stronger spatial autocorrelation in culture than genes (itself possibly a result of faster rates of cultural evolution) means that, in addition to any population boundary effects, for a given geographical scale, we expect greater between-population differences in culture than genes. If so, cultural F_{ST} or Φ_{ST} values may be particularly sensitive to the geographical scale of the population being sampled. This may help to explain why the cultural Φ_{ST} values from this study, drawn from large European language groups, and cultural F_{ST} values from countries around the globe (Bell et al., 2009) are four times larger than the cultural F_{ST} values from the considerably more localised Formosan-speaking groups (Rzeszutek et al., 2011).

Recently, empirical data on cultural and genetic F_{ST} values have been applied to debates about the units of selection in human evolution (Bell et al., 2009). The folktale variants we examine here are unlikely to affect the survival of the individuals or groups that carry them and so are essentially selectively neutral traits. Nevertheless, our findings highlight an important caveat when interpreting F_{ST} or Φ_{ST} values more generally. Bell *et al.* (2009) argue that higher cultural than genetic F_{ST} values between neighbouring groups suggests greater potential for cultural group selection. Yet, our partial Mantel tests on individual folktales show that variation is more strongly related to geographical distance (6.6% of the variation) than ethnolinguistic identity (3.7% of the variation). Hence, while populations differ and significant cultural barriers exist, geographical distance appears to be the most important factor. If this pattern generalises to other elements of culture then, because much cultural competition is likely to have played out on a local valley-to-valley or village-to-village scale, actual differences between competing groups may be much less than is indicated by F_{ST} or Φ_{ST} values calculated on the basis of large-scale ethnolinguistic identities—the same is true for genetic variation. This highlights the importance of considering the spatial dimension of cultural and genetic variation when evaluating theoretical models of competition between groups.

5.4.5 The cultural landscape of Europe

The NeighbourNet in Figure 5.3 represents graphically the pattern of regional clustering in folktale variation. The five clusters we identify provide insights into possible cultural spheres of influence in Europe since the folktale's inception. Cluster (i) includes the western European Romance-speaking populations (excluding Romanian) as well as other non-Romance-speaking western European populations (Basque, Flemish and Swiss German). Cluster (ii) includes the eastern European Slavic-speaking populations, plus other non-Slavic-speaking eastern European populations (Romanian, and Finno-Ugric speakers from Russia). Cluster (iii) includes the southeastern European populations (Turkish, Greek and Armenian). Cluster (iv) includes northern European North Germanic-speaking populations (excluding Danish), plus Finnish. Interestingly, Swedes in Finland are placed alongside Finnish, not Swedish, reinforcing the importance of geography over cultural ancestry. Cluster (v) is less obviously a geographical grouping, comprising German, Danish and Latvian in mainland northern Europe plus English, Irish and Scottish from the British Isles. The British Isles have met with waves of immigration and trade from the ancestors of these northern European groups, from Viking expansion beginning in the ninth century AD to trade networks such as the Hanseatic League, which linked the Baltic to Northern Europe and Britain from the thirteenth century AD. If this grouping is preserving the traces of early contact then the folktale stretches back beyond the earliest attested variants, which do not appear until the fourteenth century (Roberts, 1958).

5.5 Conclusion

Much has been made of analogies between processes of biological and cultural evolution and the potential for interdisciplinary cross-fertilisation (Atkinson & Gray, 2005; Bentley et al., 2004; Eerkens & Lipo, 2007; Gray et al., 2007; Mesoudi et al., 2004, 2006; O'Brien & Lyman, 2003). While there exist important disanalogies between cultural and biological processes, particularly with regard to microevolutionary transmission mechanisms (Boyd & Richerson, 1985; Henrich & Gil-White, 2001), our findings suggest that methods and theory from population genetics can nonetheless be usefully applied to characterise population structure and variation in cultural packages such as folktales. Our comparisons of the broad patterns that emerge on a continental scale in folktale and genetic diversity point to some key similarities and differences in the forces shaping the two. In addition, the location information from individual folktale variants allowed us to tease apart the relative effects of population structure and geography on cultural diversity. Future work using the approach we describe here could examine how these patterns differ across other aspects of human culture, such as variation in material culture assemblages through time in the archaeological record (Eerkens & Lipo, 2007), providing important insights into processes of cultural transmission and the interplay between human genetic and cultural evolution.

5.6 References

Atkinson, Q. D., & Gray, R. D. (2005). Curious parallels and curious connections--phylogenetic thinking in biology and historical linguistics. *Systematic Biology, 54*(4), 513-526. doi: 10.1080/10635150590950317

Barbujani, G., & Colonna, V. (2010). Human genome diversity: frequently asked questions. *Trends in Genetics, 26*(7), 285-295. doi: 10.1016/j.tig.2010.04.002

Barbujani, G., & Colonna, V. (2011). Genetic basis of human biodiversity: an update. In F. E. Zachos & J. C. Habel (Eds.), *Biodiversity hotspots: distribution and protection of conservation priority areas* (pp. 97-119). Berlin, Germany: Springer.

Barbujani, G., & Sokal, R. R. (1990). Zones of sharp genetic change in Europe are also linguistic boundaries. *Proceedings of the National Academy of Sciences, USA, 87*, 1816-1819. doi: 10.1073/pnas.87.5.1816

Bell, A. V., Richerson, P. J., & McElreath, R. (2009). Culture rather than genes provides greater scope for the evolution of large-scale human prosociality. *Proceedings of the National Academy of Sciences, USA, 106*(42), 17671-17674. doi: 10.1073/pnas.0903232106

Bentley, R. A., Hahn, M. W., & Shennan, S. J. (2004). Random drift and culture change. *Proceedings of the Royal Society B: Biological Sciences, 271*(1547), 1443-1450. doi: 10.1098/rspb.2004.2746

Borgerhoff Mulder, M., Nunn, C. L., & Towner, M. C. (2006). Cultural macroevolution and the transmission of traits. *Evolutionary Anthropology: Issues, News, and Reviews, 15*(2), 52-64. doi: 10.1002/evan.20088

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago, Il: University of Chicago Press.

Boyd, R., Richerson, P. J., Borgerhoff Mulder, M., & Durham, W. H. (1997). Are cultural phylogenies possible? In P. Weingart, P. J. Richerson, S. D. Mitchell, & S. Maasen

(Eds.), *Human by nature, between biology and the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bryant, D., Filimon, F., & Gray, R. D. (2005). Untangling our past: languages, trees, splits and networks. In R. Mace, C. J. Holden, & S. Shennan (Eds.), *The evolution of cultural diversity: phylogenetic approaches*. London, UK: UCL Press.

Bryant, D., & Moulton, V. (2004). Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution, 21*(2), 255-265. doi: 10.1093/molbev/msh018

Campbell, L., & Posner, W. J. (2008). *Language classification: history and method*. Cambridge, UK: Cambridge University Press.

Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The history and geography of human genes*. Princeton, NJ: Princeton University Press.

Currie, T. E., Greenhill, S. J., Gray, R. D., Hasegawa, T., & Mace, R. (2010). Rise and fall of political complexity in island South-East Asia and the Pacific. *Nature, 467*(7317), 801-804. doi: 10.1038/nature09461

Eerkens, J. W., & Lipo, C. P. (2005). Cultural transmission, copying errors, and the generation of variation in material culture and the archaeological record. *Journal of Anthropological Archaeology, 24*(4), 316-334. doi: 10.1016/j.jaa.2005.08.001

Eerkens, J. W., & Lipo, C. P. (2007). Cultural transmission theory and the archaeological record: providing context to understanding variation and temporal changes in material culture. *Journal of Archaeological Research, 15*(3), 239-274. doi: 10.1007/s10814-007-9013-z

Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online, 1*, 47-50.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysing molecular variance inferred from metric distance among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics, 131*, 479-491.

Fortunato, L., Holden, C. J., & Mace, R. (2006). From bridewealth to dowry? A Bayesian estimation of ancestral states of marriage transfer in Indo-European groups. *Human Nature, 17*(4), 355-376. doi: 10.1007/s12110-006-1000-4

Gachter, S., Herrmann, B., & Thoni, C. (2010). Culture and cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1553), 2651-2661. doi: 10.1098/rstb.2010.0135

Gillies, W. (1993). Scottish Gaelic. In M. J. Ball (Ed.), *The Celtic languages* (pp. 145-227). London, UK: Routledge.

Gray, R. D., & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature, 426*(6965), 435-439. doi: 10.1038/nature02138

Gray, R. D., Bryant, D., & Greenhill, S. J. (2010). On the shape and fabric of human history. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1559), 3923-3933. doi: 10.1098/rstb.2010.0162

Gray, R. D., Greenhill, S., & Ross, R. M. (2007). The pleasures and perils of Darwinizing culture (using phylogenies). *Biological Theory, 2*(4), 360-375. doi: 10.1162/biot.2007.2.4.360

Handley, L. J. L., Manica, A., Goudet, J., & Balloux, F. (2007). Going the distance: human population genetics in a clinal world. *Trends in Genetics, 23*(9), 432-439. doi: 10.1016/j.tig.2007.07.002

Henrich, J., Boyd, R., McElreath, R., Gurven, M., Richerson, P. J., Ensminger, J., . . . Ziker, J. (2012). Culture does account for variation in game behavior. *Proceedings of the National Academy of Sciences, USA, 109*(2), E32-E33. doi: 10.1073/pnas.1118607109

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior, 22*, 165-196. doi: 10.1016/s1090-5138(00)00071-4

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nature Reviews Genetics, 10*(9), 639-650. doi: 10.1038/nrg2611

Itan, Y., Powell, A., Beaumont, M. A., Burger, J., & Thomas, M. G. (2009). The origins of lactase persistence in Europe. *PLoS Computation Biology, 5*(8), e1000491. doi: 10.1371/journal.pcbi.1000491

Jordan, F. M., Gray, R. D., Greenhill, S. J., & Mace, R. (2009). Matrilocal residence is ancestral in Austronesian societies. *Proceedings of the Royal Society B: Biological Sciences, 276*(1664), 1957-1964. doi: 10.1098/rspb.2009.0088

Kempe, M., Lycett, S., & Mesoudi, A. (2012). An experimental test of the accumulated copying error model of cultural mutation for Acheulean handaxe size. *PLOS ONE, 7*(11), e48333. doi: 10.1371/journal.pone.0048333

Kimura, M. (1983). *The neutral rheory of molecular evolution*. Cambridge, UK: Cambridge University Press.

Lamba, S., & Mace, R. (2012). Demography and ecology drive variation in cooperation across human populations. *Proceedings of the National Academy of Sciences, USA, 108*(35), 14426-14430. doi: 10.1073/pnas.1105186108

Lao, O., Lu, T. T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., . . . Kayser, M. (2008). Correlation between genetic and geographic structure in Europe. *Current Biology, 18*(16), 1241-1248. doi: 10.1016/j.cub.2008.07.049

Lewontin, R. C. (1972). The apportionment of human diversity. *Evolutionary Biology, 6*, 381-398. doi: 10.1007/978-1-4684-9063-3_14

Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., . . . Myers, R. M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science, 319*(5866), 1100-1104. doi: 10.3410/f.1104172.560167

Lycett, S. J. (2008). Acheulean variation and selection: does handaxe symmetry fit neutral expectations? *Journal of Archaeological Science, 35*(9), 2640-2648. doi: 10.1016/j.jas.2008.05.002

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research, 27*(2), 209-220.

Marcantonio, A. (2002). *The Uralic language family: facts, myths and statistics*. Oxford, UK: Blackwell Publishing.

Marks, J. (2010). Ten facts about human variation. In M. P. Muehleenbein (Ed.), *Human evolutionary biology* (pp. 265-276). Cambridge, UK: Cambridge University Press.

Maynard Smith, J. (1998). *Evolutionary genetics, 2nd ed*. Oxford, UK: Oxford University Press.

Mesoudi, A., Whiten, A., & Laland, K. N. (2004). Is human cultural evolution Darwinian? Evidence reviewed from the perspective of The Origin of Species. *Evolution, 58*(1), 1-11. doi: 10.1111/j.0014-3820.2004.tb01568.x

Mesoudi, A., Whiten, A., & Laland, K. N. (2006). Towards a unified science of cultural evolution. *Behavioral and Brain Sciences, 29*(4), 329-383. doi: 10.1017/s0140525x06009083

Moore, J. H. (1994). Putting anthropology back together again: the ethnogenetic critique of cladistic theory. *American Anthropologist, 96*(4), 925-948. doi: 10.1525/aa.1994.96.4.02a00110

Neiman, F. D. (1995). Stylistic variation in evolutionary perspective: inferences from decroative diversity and interassemblage distance in Illinois woodland ceramic assemblages. *American Antiquity, 60*(1), 7-36. doi: 10.2307/282074

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., . . . Bustamante, C. D. (2008). Genes mirror geography within Europe. *Nature, 456*(7218), 98-101. doi: 10.1038/nature07331

Novembre, J., & Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics, 40*(5), 646-649. doi: 10.1038/ng.139

O'Brien, M. J., & Lyman, R. L. (2003). *Cladistics and archaeology*. Salt Lake City, Ut: University of Utah Press.

Pagel, M. (2009). Human language as a culturally transmitted replicator. *Nature Reviews Genetics, 10*(6), 405-415. doi: 10.1038/nrg2560

Pagel, M., & Mace, R. (2004). The cultural wealth of nations. *Nature, 428*(6980), 275-278. doi: 10.1038/428275a

Peakall, R., Ruibal, M., & Lindenmayer, D. B. (2003). Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, Rattus fuscipes. *Evolution, 57*(5), 1182-1195. doi: 10.1111/J.0014-3820.2003.Tb00327.X

Peakall, R., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes, 6*(1), 288-295. doi: 10.1111/J.1471-8286.2005.01155.X

Roberts, W. E. (1958). *The tale of the kind and the unkind girls: Aa-Th 480 and related tales*. Germany: De Gruyter.

Rogers, D. S., & Ehrlich, P. R. (2008). Natural selection and cultural rates of change. *Proceedings of the National Academy of Sciences, USA, 105*(9), 3416-3420. doi: 10.1073/pnas.0711802105

Rosser, Z. H., Zerjal, T., Hurles, M. E., Adojaan, M., Alavantic, D., Amorim, A., . . . Jobling, M. A. (2000). Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *American Journal of Human Genetics, 67*(6), 1526-1543. doi: 10.1086/316890

Rzeszutek, T., Savage, P. E., & Brown, S. (2011). The structure of cross-cultural musical diversity. *Proceedings of the Royal Society B: Biological Sciences*. doi: 10.1098/rspb.2011.1750

Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Multiple-regression and correlation extensions of the Mantel Test of matrix correspondence. *Systematic Zoology, 35*(4), 627-632. doi: 10.2307/2413122

Thompson, S. (1946). *The folktale*. New York, NY: Holt, Rinehart and Winston, Inc.

Uther, H. J. (2004). *The types of international folktales: classification and bibliography, based on the system of Antti Aarne and Stith Thompson*. Helsinki, Finland: Suomalainen Tiedeakatemia, Academia Scientiarum Fennica.

Wright, S. (1943). Isolation by distance. *Genetics, 28*(114-138).

Chapter Six: Cultural differentiation does not entail group-level structure: The case for geographically explicit analysis

6.1 Abstract

Richerson *et al.* (in press) argue that relatively large cultural F_{ST} values provide evidence for group structure and therefore scope for group selection. However, recent research on spatial patterns of cultural variation demonstrates that, as in the genetic case, apparent group structure can be a consequence of geographic clines, not group barriers. Such a pattern limits the scope for cultural group selection.

6.2 Commentary

Richerson *et al.* (in press) present a wide-ranging synthesis of evidence for cultural group selection. An innovative feature of their argument is that they draw attention to the importance of quantifying the apportionment of cultural variation within and between groups. They calculate F_{ST} values for 29 aspects of culture between neighbouring groups and argue that relatively large cultural F_{ST} values provide scope for cultural group selection. In addition, they cite Bell, Richerson and McElreath's (2009) finding that average cultural F_{ST} values (as indexed by responses to the World Values Survey) are greater for neighbouring groups than average genetic F_{ST} values as evidence that there is greater potential for cultural group selection than genetic group selection. While we agree that this framework for quantifying cultural variation holds considerable potential, we challenge Richerson *et al.*'s interpretation of the existing empirical data.

The F_{ST} statistic belongs to a broad family of "fixation indices"—statistics developed by population geneticists to study genetic differentiation between populations (Holsinger & Weir, 2009). Several studies have demonstrated that human genetic variation is predominantly clinal, with differentiation between populations being strongly predicted by geographic distance across a variety of biological markers (Handley, Manica, Goudet, & Balloux, 2007). Discontinuities do exist (typically associated with geographic obstacles), but many apparent genetic barriers have proven to be artifacts of heterogeneous spatial sampling (Handley et al., 2007). A clinal pattern of variation is consistent with an "isolation by distance" (IBD) model (Wright, 1943), in which individuals tend to migrate limited distances relative to the total geographic range of the species. Under IBD, a gradual blending of one population into the next is predicted, rather than group barriers. Nevertheless, two sampling locations can produce significant F_{ST} values simply due to their geographic separation.

None of the 29 cultural F_{ST} estimates reported by Richerson *et al.* nor the cultural F_{ST} estimates reported by Bell *et al.* (2009) have been analysed within a spatially explicit framework. This renders these estimates difficult to interpret, because, as in the genetic case, apparent population structure could be an artifact of discontinuous spatial sampling, rather than cultural barriers.

Recently, we published a study that quantified the extent to which geography and group affiliation independently predict cultural differentiation between ethnolinguistic groups (Ross, Greenhill, & Atkinson, 2013). We used geographic coordinates and coded narrative elements of 700 versions of a folktale from 31 European groups, analysing both individual folktales and group level differentiation using Φ_{ST}, a fixation index that is closely related to F_{ST} (Holsinger & Weir, 2009). We found significant differentiation between groups with an average Φ_{ST} of 0.091, indicating that, on average, 9.1% of the variation between individual folktales was between groups, which is considerably higher than the genetic differentiation found between comparable European populations (Lao et al., 2008; Novembre et al., 2008). However, incorporating geography into the analysis revealed that at the level of individual folktales, geographic distance explains considerably more variation between folktales than group boundaries (6.6% of variance versus 3.7%). Such a pattern of predominantly clinal variation is consistent with IBD-like cultural diffusion processes. This means that geographically close individuals/tales from neighbouring groups tend to be more similar than is suggested by the relatively large average cultural Φ_{ST}, thereby limiting the scope for cultural group selection.

Two recent studies speak to the generality of our findings and the F_{ST} estimates from Richerson *et al.* First, Brown *et al.* (2014) examined group-level folk song differentiation among nine indigenous ethnolinguistic groups in Taiwan. They found significant cultural differentiation between groups. Intriguingly, however, cultural Φ_{ST} was found to be an order of magnitude *smaller* than genetic Φ_{ST} for the same groups, contra Bell *et al.*'s (2009) hypothesis that human groups generally show a higher degree of cultural differentiation than genetic differentiation. This result is particularly pertinent to discussions about human evolution, because the folk song data are drawn from small-scale indigenous societies whose lifestyles and group structure better approximate those of our ancestors than the large-scale multiethnic nation states studied by Bell *et al.* (2009). Second, Shennan, Crema and Kerig (2015) examined individual level variation and group-level differentiation in two material culture complexes – pottery and personal ornaments – from 361 sites of 22 putative Neolithic cultural groups in Europe. At the level of individual artefacts, cultural affiliation was an independent predictor of pottery variation, while geography was not; but both cultural affiliation and geography were independent predictors of ornament variation. At the group-level, they found significant cultural differentiation, with geography predicting differentiation in ornaments but not pottery. This result suggests that the relative influence of cultural barriers and geographic effects can vary across different cultural markers in comparable populations, just as is the case for different genetic markers (i.e., autosomal DNA, mitochondrial DNA, and Y-chromosome; Jobling, 2012). Caution is therefore warranted in making generalisations about the relationships between groups on the basis of analysis of a limited range of cultural traits.

While we support the rigorous analysis of empirical data to quantify cultural variation, the evidence we present here suggests that cultural differentiation between groups varies considerably across cultural domains and spatial scales and is often best explained in terms of geographic clines – a pattern that suggests IBD-like cultural processes and limited scope for cultural group selection between neighbours. We note that, in the absence of stable individual level trait differences between neighbouring groups, the most important forms of variation for cultural group selection may be group level traits (Smaldino, 2014), such as the presence or absence of particular religious or political institutions, rather than the values, stories, songs, or material possessions of individuals. Further work that examines individual variation and group-level differentiation using a geographically explicit framework across the full spectrum of aspects of human culture is needed.

6.3 References

Bell, A. V., Richerson, P. J., & McElreath, R. (2009). Culture rather than genes provides greater scope for the evolution of large-scale human prosociality. *Proceedings of the National Academy of Sciences USA, 106*(42), 17671-17674. doi: 10.1073/pnas.0903232106

Brown, S., Savage, P. E., Ko, A. M., Stoneking, M., Ko, Y. C., Loo, J. H., & Trejaut, J. A. (2014). Correlations in the population structure of music, genes and language. *Proceedings of the Royal Society B: Biological Sciences, 281*(1774), 20132072. doi: 10.1098/rspb.2013.2072

Handley, L. J. L., Manica, A., Goudet, J., & Balloux, F. (2007). Going the distance: human population genetics in a clinal world. *Trends in Genetics, 23*(9), 432-439. doi: 10.1016/j.tig.2007.07.002

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nature Reviews Genetics, 10*(9), 639-650. doi: 10.1038/nrg2611

Jobling, M. A. (2012). The impact of recent events on human genetic diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1590), 793-799. doi: 10.1098/rstb.2011.0297

Lao, O., Lu, T. T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., . . . Kayser, M. (2008). Correlation between Genetic and Geographic Structure in Europe. *Current Biology, 18*(16), 1241-1248. doi: 10.1016/j.cub.2008.07.049

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., . . . Bustamante, C. D. (2008). Genes mirror geography within Europe. *Nature, 456*(7218), 98-101. doi: 10.1038/nature07331

Richerson, P., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., . . . Zefferman, M. R. (in press). Cultural group selection plays an essential role in explaining human cooperation: a sketch of the evidence. *Behavioral and Brain Sciences, Advance online publication*. doi: 10.1017/s0140525x1400106x

Ross, R. M., Greenhill, S. J., & Atkinson, Q. D. (2013). Population structure and cultural geography of a folktale in Europe. *Proceedings of the Royal Society B: Biological Sciences, 280*(1756), 20123065. doi: 10.1098/rspb.2012.3065

Shennan, S. J., Crema, E. R., & Kerig, T. (2015). Isolation-by-distance, homophily, and “core” vs. “package” cultural evolution models in Neolithic Europe. *Evolution and Human Behavior*, 36(2):103-119. doi: 10.1016/j.evolhumbehav.2014.09.006

Smaldino, P. E. (2014). The cultural evolution of emergent group-level traits. *Behavioral and Brain Sciences, 37*(3), 243-254. doi: 10.1017/S0140525X13001544

Wright, S. (1943). Isolation by distance. *Genetics, 28*, 114-138.

Chapter Seven: Folktale transmission in the Arctic provides evidence for high bandwidth social learning among hunter-gatherer groups

7.1 Abstract

There exist striking resemblances in the stories of ethnolinguistic groups separated by vast geographic distances, with nearby groups having the most in common. The causes of these geographic associations are uncertain. Here we use method and theory from population genetics to examine cultural transmission in folktale inventories of 18 hunter-gatherer groups spread across 6000 km of Siberia, Alaska, Canada, and Greenland. We find that linguistic relatedness and geographic proximity independently predict overlap in folktale inventories, which provides evidence for both vertical transmission down cultural lineages and horizontal transmission between groups. These results suggest that high-bandwidth social learning across group boundaries is a feature of traditional hunter-gatherers, which may help explain how complex cultural traditions can develop and be retained in ostensibly small groups.

7.2 Introduction

It has long been recognised that nearby "ethnolinguistic groups" (groups defined by shared language and cultural identity; henceforth "groups") share more cultural traits than those that are far away (Boas, 1896). In recent years, method and theory from evolutionary biology has been used to disentangle the cultural evolutionary processes that account for this geographic gradient (Borgerhoff Mulder et al., 2006; Gray et al., 2007; Levinson & Gray, 2012; Mace & Jordan, 2011; Nettle, 2009). Three broad classes of processes have been identified. First, vertical transmission—new groups inherit cultural traits from a parent group, and groups that have diverged recently have had less time to spread. Second, horizontal transmission—contemporaneous groups exchange cultural traits, and groups that are near one another have had more opportunities for cultural exchange. Third, independent innovation—groups that are near one another tend to live in similar environments, and groups can converge on similar cultural traits in response to shared ecological opportunities and challenges. The relative influence of these three processes varies across aspects of culture. Language frequently exhibits highly tree-like patterns of vertical transmission (Bouckaert et al., 2012; Gray et al., 2009; Lee & Hasegawa, 2011), as do a variety of social practices (Currie et al., 2010; F. M. Jordan et al., 2009; Opie et al., 2014), and aspects of material culture (P. Jordan & Shennan, 2009; Larsen, 2011; Tehrani & Collard, 2002). By contrast, other social practices and aspects of material culture show high levels of horizontal transmission (Gray et al., 2010; P. Jordan & Shennan, 2003; Towner et al., 2012). And environmental commonalities can spur independent cultural innovation (Saslis-Lagoudakis et al., 2014).

Storytelling is a highly conspicuous cross-cultural universal (D. E. Brown, 1991), and a variety of theories propose that it has played a key role in human evolution (B. Boyd, 2009; Carroll, 2006; Coe et al., 2006; Davies, 2012; Gottschall, 2012; Scalise Sugiyama, 2001; Wiessner, 2014). The importance of cultural evolutionary processes is underscored by the fact that the folktales, myths and legends of nearby groups tend to be more alike than those of groups separated by large distances. Folktales, in particular, have been the target of a long tradition of careful documentation and classification (Goldberg, 1984; Krohn, 1926; Uther, 1997, 2004). Nevertheless, the processes underlying this geographic relationship remain poorly understood, with competing theories invoking vertical transmission, horizontal transmission, and independent innovation (Dorson, 1972; Dundes, 1986; Teverson, 2013; Thompson, 1946).

Research on folktale transmission has the potential to inform broad debates about cultural evolution (Ross & Atkinson, in press; Ross et al., 2013; Tehrani, 2013), including recent discussions about the relationship between cultural complexity and demography. It has been hypothesised that a suitably large population of potential teachers is crucial for developing and maintaining complex culture, with larger groups predicted to have richer and more complex cultural repertoires than smaller groups (Henrich, 2004). To date, research on this broad hypothesis has focused on the specific case of toolkit complexity, particularly food-getting technologies (Collard et al., 2013a; Collard et al., 2013b; Henrich, 2004; Kline & Boyd, 2010; Powell et al., 2009; Read, 2006, 2012). One challenge for interpreting toolkit data is to account for how some small groups developed and retained remarkably complex toolkits despite their small group sizes. One possibility is that the "effective population size" (the size of the interacting pool of social learners) of groups can be substantially larger than the size of groups themselves if social learning occurs across group boundaries (Henrich, 2004). This is an important question for empirical examination. However, it is difficult to index the degree of intimacy of inter-group social learning by analysing toolkits because tools can diffuse between groups not only through explicit teaching and apprenticeship (Sterelny, 2012; i.e. high intimacy), but also through trading and stealing (i.e. low intimacy). We propose that horizontal transmission of folktales provides a novel marker of high intimacy, high bandwidth social learning between groups since folktales, unlike food getting technologies, are necessarily exchanged through language.

Although cultural evolution approaches have been widely applied to the study of material culture, language, and social practices (Mesoudi, 2011), little research has explored storytelling. Recently, however, Ross *et al.* (2013) used method and theory from population genetics to quantify the extent to which geographic distance, group affiliation, and cultural ancestry independently predict the distribution of 700 variants of a folktale across 31 European groups. At the level of individual folktales, they found that geography explained most variation, followed by group affiliation, and finally cultural ancestry. At the group level, they found significant cultural differentiation among groups, with geography explaining more differentiation than is explained by cultural ancestry. Tehrani (2013) subsequently used phylogenetic methods to examine 58 variants of a folktale from cultures of Europe, Africa and East Asia. He found evidence for phylogenetic signal, with the degree of branching and blending varying in different geographic regions.

Many of the groups included in the studies by Ross *et al.* (2013) and Tehrani (2013) are characterised by social structures, cultural practices, population densities, and technologies that are radically different to those of the hunter-gatherer groups that characterised much of our recent evolutionary past. In particular, most of the groups are large, complex nation states with writing systems, print technologies, and widespread literacy and schooling. Importantly, it has been argued that these cultural innovations initiate profound changes in patterns and processes of cultural transmission, malleability and fixity of traditions, group identity, cooperation, memory, and other cognitive processes (Eisenstein, 1979; Goody & Watt, 1963; Mullins et al., 2013; Ong, 1982; Poe, 2010; Rubin, 1995). Furthermore, these cultural innovations feature prominently in recent debates about folktale transmission, particularly in Europe. Some scholars maintain that many of the earliest attested European folktales have long been transmitted orally and were widely distributed many centuries before being written down (Zipes, 2006, 2012). By contrast, other scholars argue that many of these folktales were invented relatively recently and became widely distributed primarily as a result of dissemination via printed media (Bottigheimer, 2009; de Blécourt, 2012). Consequently, it is uncertain whether it is appropriate to use the results of the studies by Ross *et al.* (2013) and Tehrani (2013) to make inferences about cultural transmission in traditional hunter-gatherer groups.

In the present study we quantify the extent to which geographic distance and cultural ancestry predict overlap in the folktale inventories of 18 groups spread across vast Arctic regions of Siberia, Alaska, Canada, and Greenland. This dataset is especially well-suited to enhance our understanding of two key issues about cultural evolution. First, in contrast to the groups examined in recent studies of folktale evolution (Ross et al., 2013; Tehrani, 2013), these Arctic groups were hunter-gatherers who lacked writing systems prior to European colonisation. Consequently, these groups are particularly useful for making inferences about patterns of cultural transmission prior to recent cultural and technological developments. Second, many of these Arctic groups were very small, yet had strikingly complex toolkits (R. Boyd et al., 2011). Consequently, evidence for horizontal transmission of folktales between these groups would suggest that the effective population for high intimacy, high bandwidth cultural learning is larger than nominal group size, and thus provide scope for explaining how complex cultural traditions can develop and be retained in ostensibly small groups.

7.3 Materials and methods

7.3.1 Data

We sourced folktale data from a study of Artic folktales (Sheppard, 1998)[10]. This study coded the presence or absence of 45 folktales across 18 groups from Arctic regions of Siberia, Alaska, Canada, and Greenland. In two instances, the presence of a folktale in a group was

[10] The data presented in Sheppard (1998) include a number of inconsistencies. Where these inconsistences could not be harmonised we favoured the data presented in the appendix, rather than the table in the main text, as per the recommendation of a friend and colleague of the deceased author who has some familiarity with the raw data (Kenneth L. Pratt, pers. com. 28 Nov. 2013).

coded as “possible fragment,” which we recoded as absent prior to analysis. The mean number of folktales per group was 17.72 (Baffin, 21; Bering Strait, 23; Caribou, 22; Chugach, 7; Chukchi, 15; Copper, 22; Greenland, 27; Iglulik, 25; Koniag, 12; Labrador, 14; Mackenzie, 14; Mainland Southwest Alaska, 15; Netsilik, 27; North Alaska, 22; Northwest Alaska, 28; Nunivak Island, 5; Quebec, 13; Siberian Yupik, 9).

We recoded the presence or absence of a folktale as “1” or “0”, respectively, to produce a matrix of 45 binary traits across 18 groups. For analysis, this presence/absence matrix was converted to a pairwise folktale distance matrix using Jaccard distances. The Jaccard distance for each pair of groups was calculated as the sum of the number of folktales that are present in one group but not the other, divided by the sum of the number of folktales that are present in one or both groups. We used the Jaccard distance because it standardises for the number of traits observed for each pair and shared absences do not contribute to similarity, which is particularly appropriate for cultural data (Rogers & Ehrlich, 2008).

We estimated geographic coordinates for groups using information provided by Sheppard (1998) and sources referenced therein. These geographic coordinates were used to calculate a pairwise geographic distance matrix. Pairwise distances were calculated using great circle distances in GenAlEx v6.501 (Peakall & Smouse, 2006, 2012).

We used language divergence to index cultural ancestry. The degree of divergence between languages was inferred using the language classifications presented in *Ethnologue* (Lewis, 2013). All but one of the groups included in the dataset speak a language that is a member of the Eskimo-Aleut language family. The other language, Chukchi, is a member of the small Chukotko-Kamchatkan language family. Pairwise language distances were coded as follows: 1 – closely related dialects, 2 – same shallow branch of the same language family, 3 – same deep branch of the same language family, 4 – different deep branches of the same language family, 5 – different language families. These pairwise distances were used to infer a pairwise language distance matrix.

7.3.2 Analysis

We quantified the association between geographic, language, and folktale distance matrices by calculating correlations and partial correlations using Mantel and partial Mantel tests (Mantel, 1967; Smouse et al., 1986) in Arlequin v3.5.1.3 (Excoffier & Lischer, 2010). Statistical significance was assessed using 10,000 random permutations.

We calculated spatial autocorrelations for folktale distances using the method implemented in GenAlEx v6.501 (Banks & Peakall, 2012). This method was used to calculate an autocorrelation coefficient *r* within 500 km distance classes for the folktale distance matrix, which provides a measure of the similarity between pairs of folktales whose geographic separation falls within each distance class. Tests for statistical significance were performed using two methods: calculating *r* across 10,000 random permutations and 10,000 bootstrap estimates (Peakall et al., 2003).

We performed Principle Coordinate Analysis (PCoA; also known as Multidimensional Scaling) to visualise major trends in the data using the covariance-standardised method implemented in GenAlEx v6.501 (Peakall & Smouse, 2006, 2012). The first two dimensions of the PCoA were correlated against latitude and longitude to explore geographic associations in the data.

In order to further visualise specific relationships among groups, we constructed a NeighbourNet (Bryant & Moulton, 2004) using the folktale distance matrix in SplitsTree v4.13.1 (Huson & Bryant, 2006). The NeighbourNet algorithm is useful for identifying complex transmission histories of group divergence and convergence (Bryant et al., 2005). The NeighbourNet algorithm does not assume a strictly tree-like model of evolution; instead, evidence for such a model appears as bifurcating “tree-like” splits in the graph. Conversely, evidence for independent innovation or horizontal transmission appears as reticulate, “box-like” structures representing conflicting subdivisions.

7.4 Results and discussion

7.4.1 Horizontal transmission

The present results provide strong evidence that between-group social learning plays an important role in the evolution of folktales in these groups. A Mantel test provides strong evidence for a moderate correlation between folktale similarity and geography ($r = .506$, $p < .001$), which suggests that the folktale inventories of nearby groups are more similar than those of distant groups. Crucially, a partial Mantel test provides strong evidence for a moderate correlation after controlling for language ($r = .409$, $p < .001$), which supports the hypothesis that there is an independent effect of horizontal transmission after controlling for vertical transmission down cultural lineages.

Spatial autocorrelation analysis reveals an overall spatial autocorrelation (omega = 129.975, $p < .001$). The positive autocorrelation is of greatest magnitude at the shortest distance class of 500 km and decreases in magnitude as distances increase until passing through the intercept at approximately 2403 km and levelling off for longer distance classes (see Figure 7.1).

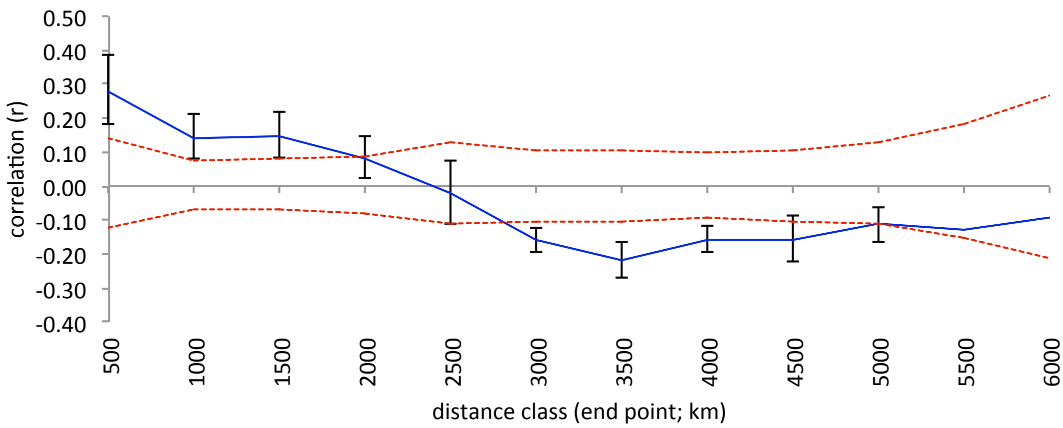


Figure 7.1 Folktale spatial autocorrelation analysis showing correlation coefficient (*r*) as a function of geographic distance. The permuted 95% confidence interval (dashed lines) and the bootstrap 95% confidence interval error bars are also shown.

PCoA of the folktale matrix reveals that the first axis of variation accounts for 21.66% of the variance (eigenvalue = 0.314) and the second axis accounts for 10.49% of the variance (eigenvalue = 0.152). Pearson product-moment correlations reveal that the first axis of variation is correlated with longitude ($r = .924$, $p < .001$), but not latitude ($r = .002$, $p = .995$); and the second axis of variation is correlated with latitude ($r = .610$, $p < .001$), but not longitude ($r = .040$, $p = .876$). To visualise the relationship between geography and the first axis of variation we plotted the results of the ordinary Kriging spatial interpolation technique implemented in ArcGIS v10.2 (ESRI, 2011) on a map (see Figure 7.2). This map suggests a smooth gradient across the major geographic axis of dispersion of these groups. Nevertheless, given the relatively small number of data points and large geographic distances between them, some caution is warranted when interpreting this apparently smooth cline since it is likely to be difficult to identify barriers if they exist.

Figure 7.2 Approximate geographic locations of the 18 groups. Colouring shows a simple Kriging interpolation of the first axis of variation of Principle Coordinate Analysis (PCoA).

This clear geographic patterning, which cannot be explained by vertical transmission down cultural lineages, is consistent with evidence that many Arctic groups were embedded in complex systems of interaction, including trade networks, friendly relations, and hostile relations (Aporta, 2009; Burch Jr., 2005; Friesen, 2013; Pratt, 2012). Our folktale analyses add a new dimension to this scholarship by providing quantitative evidence for the existence of between group cultural learning of sufficiently high bandwidth and fidelity for orally transmitted folktales to diffuse between groups. Consequently, our result provides novel evidence that effective population size can be larger than group size, even for complex orally transmitted culture, an important finding for theories that link cultural complexity to demography (Henrich, 2004; Kline & Boyd, 2010; Powell et al., 2009).

7.4.2 Vertical transmission

The present results also provide strong evidence that cultural ancestry plays an important role in the evolution of folktales in these groups. A Mantel test provides strong evidence for a moderate correlation between folktale similarity and language relatedness ($r = .439$, $p < .001$), which suggests that the folktale inventories of groups that diverged recently are more similar than those of groups that diverged less recently. Crucially, a partial Mantel test provides strong evidence for a moderate correlation after controlling for geography ($r = .310$, $p = .007$), which supports the hypothesis that there is an independent effect of vertical transmission down language lineages after controlling for horizontal transmission between groups.

The Eskimo-Aleut language family is thought to have originated in coastal Alaska up to 5000 years ago (Fortescue, 1998, 2013). Archaeological and linguistic evidence suggests that starting at about 1000 AD there was rapid migration eastward into Canada and Greenland (Fortescue, 1998, 2013). This "Thule migration" is thought to have resulted in the replacement of earlier Dorset peoples, seemingly with little or no inbreeding (Raghavan, 2014), exchange of material culture (Friesen, 2013), or influence on Eskimo-Aleut languages (Fortescue, 2013). That a major expansion in the Eskimo branch of this language family occurred relatively recently might help account for our strong evidence for vertical transmission of orally transmitted folktales.

The NeighbourNet (Figure 7.3) reveals that the vertical transmission demonstrated using Mantel tests is not very tree-like, with the box-like structures in the NeighbourNet providing evidence for widespread reticulation. This can be interpreted as providing evidence that these folktales were not transmitted as cohesive cultural packages but each have disparate histories. Such an interpretation supports Sheppard's (1998) suggestion that the folktales may have originated at diverse locations. Interestingly, large splits in the NeighbourNet point to two geographic clusters: a western cluster and an eastern cluster, with Copper, the group that is geographically situated between these two clusters, falling between these clusters on the NeighbourNet. These large splits might be the result of geographic or cultural barriers to the spread of folktales between these broad geographic regions. However, it is also possible that these apparent barriers are artifacts of the low spatial density sampling between these clusters.

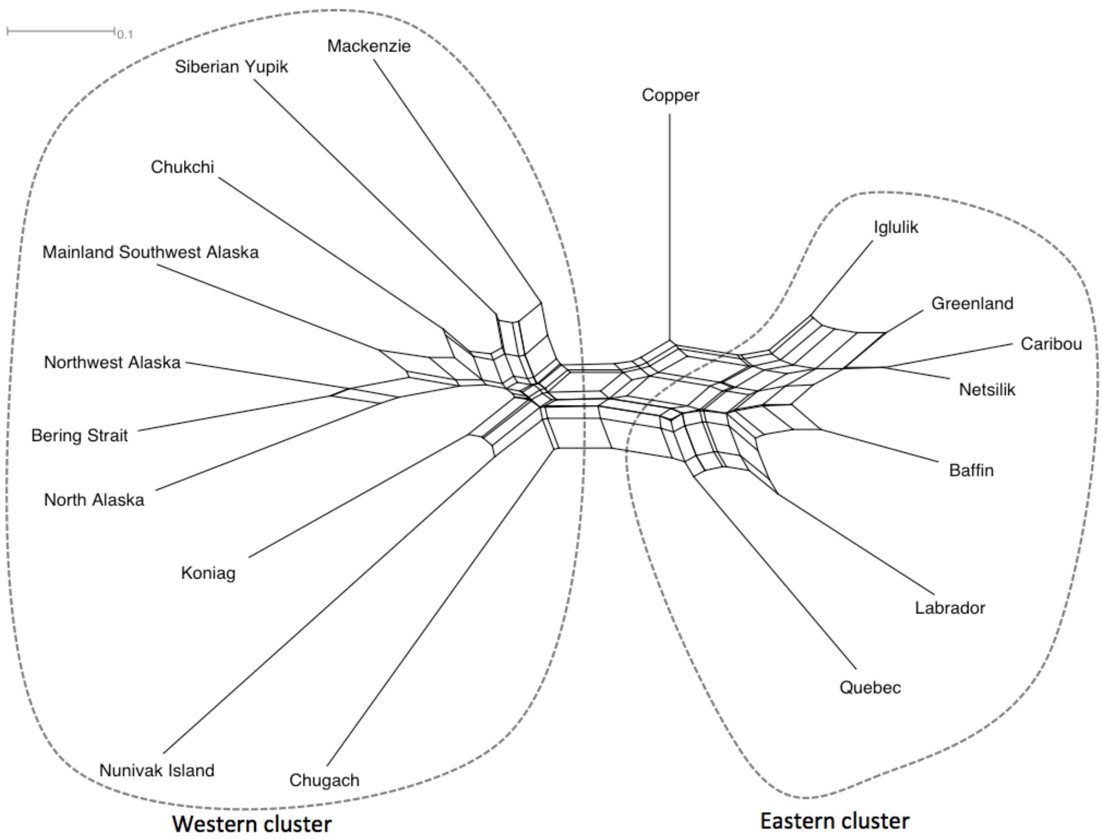


Figure 7.3 NeighbourNet of the 18 groups based on the Jaccard folktale distance matrix.

7.4.3 Comparisons with other folktale analyses

The associations found in the present analyses are somewhat stronger than those found in a comparable study of European folktale variation. Ross *et al.*'s (2013) analysis of 700 individual variants of a folktale in 31 groups in Europe (in terms of group level differentiation using the Φ_{ST} statistic) revealed a relatively strong correlation between folktale differentiation and geography ($r = 0.384$, $p < .001$), even when controlling for language relatedness ($r = 0.355$, $p = .005$). They also found a correlation between folktale differentiation and language relatedness ($r = 0.274$, $p = .014$), but no correlation when controlling for geography ($r = 0.207$, $p = .099$). It is possible that the stronger associations found in the present study are due to differences in patterns of transmission in Europe and the Arctic. One intriguing possibility is that the development of writing, print technology, and widespread literacy weakened the association between folktales, geography, and cultural ancestry in Europe, which would be consistent with theories that argue that these cultural innovations precipitate major changes in cultural transmission (Eisenstein, 1979; Goody & Watt, 1963; Mullins et al., 2013; Ong, 1982; Poe, 2010; Rubin, 1995), including folktale transmission (Bottigheimer, 2009; de Blécourt, 2012). Nevertheless, there exist important differences between the two studies with respect to how the folktale data were coded. Of particular note, the study of European folktales examined variation in narrative elements in *related* folktales, while the present study examined the presence or absence of *different* folktales. Consequently, it is possible that differences in the "grain of analysis" (Godfrey-Smith, 2012) of these datasets might be responsible for variation in the strength of associations.

7.4.4 Limitations and future research

There are at least four respects in which the data analysed in the present study are somewhat imprecise. First, traditional linguistic subgroup classifications were used to index cultural ancestry. These classifications are not ideal estimates of relative time depths since a variety of factors, such as branching (Atkinson et al., 2008) and group size (Bromhan et al., 2015), can influence rates of language evolution. Second, many Arctic groups were highly mobile and traversed large geographic ranges, making the estimated locations for some groups inexact. Third, the number of folktales examined by Sheppard (1998) varies considerably across groups. Although the primary reason for this is likely to be differences in the true size of folktale inventories of different groups, it is likely that some of this variation is also due to some groups having been documented more thoroughly than others. Fourth, members of some of these groups had become literate in their own language before folktale data were collected. For instance, the Greenland and Labrador Inuit groups became literate during the 18th century, and other Inuit groups in Alaska and Canada followed during the 19th century (Dorais, 2010). Nevertheless, we are not aware of any evidence to suggest that writing was used to transmit folktales between groups.

Despite these limitations, we suggest that our results are likely to reflect genuine patterns of cultural transmission among these groups. The associations between folktales, geography, and cultural ancestry are of moderate size and strongly supported by statistical tests. Furthermore, we have no reason to believe that imprecisions are manifestations of underlying systemic biases that might inflate the sizes of associations. Nevertheless, we suggest that future research that enhances the precision of the data could prove to be useful. In particular, the source materials used by Sheppard (1998) could be used to code the presence and absence of narrative elements in different versions of the same folktale, and thus could be used to examine these folktales at the "grain of analysis" used in studies that focused on European groups (Ross et al., 2013; Tehrani, 2013), which provides useful data for testing theories about cultural group selection (Ross & Atkinson, in press). In addition, Eskimo-Aleut cognate data (e.g. Fortescue et al., 2011) could be used to infer an Eskimo-Aleut language phylogeny, and thus index cultural ancestry with greater precision, as has been done in studies of other language families (Currie et al., 2010; F. M. Jordan et al., 2009; Opie et al., 2014). Furthermore, other cultural data could be collected and analysed to provide a more comprehensive picture of the relative importance of vertical transmission and horizontal transmission in these groups; tools (especially food-getting technologies) would be particularly appropriate since they have been the focus of much research and discussion (Collard et al., 2013a; Collard et al., 2013b; Henrich, 2004; Kline & Boyd, 2010; Powell et al., 2009; Read, 2006, 2012). Likewise, it would be useful to examine associations between cultural data and genetic data for the same groups, as has been done in a recent study of indigenous Taiwanese folk music (S. Brown et al., 2014). Such follow up studies could make important contributions to general debates about the extent to which genes, culture, and language travel together; and specific debates about the peopling of the Americas (Fortescue, 1998, 2013; Friesen, 2013; Raghavan, 2014).

7.5 Conclusion

The present study makes a novel contribution to our understanding of cultural transmission. Previous studies of folktales focused on large nation states with writing systems, print technologies, and high rates of literacy (Ross et al., 2013; Tehrani, 2013) for which there is vigorous debate about the relative importance of written and oral transmission of folktales (Bottigheimer, 2009; de Blécourt, 2012; Zipes, 2006, 2012). The present results demonstrate that in traditional hunter-gather groups without writing systems both vertical transmission down cultural lineages and horizontal transmission between groups play important roles in folktale transmission. In addition, these results provide a novel source of quantitative evidence for high bandwidth, high intimacy social learning across group boundaries. As a consequence, these results provide evidence that the effective population size for social learning of orally transmitted culture can be larger than nominal group size, which might help explain how cumulative cultural evolution is possible in small groups.

7.6 References

Aporta, C. (2009). The trail as home: Inuit and their pan-Arctic network of routes. *Human Ecology, 37*(2), 131-146. doi: 10.1007/s10745-009-9213-x

Atkinson, Q. D., Meade, A., Venditti, C., Greenhill, S. J., & Pagel, M. (2008). Language evolves in punctuational bursts. *Science, 319*(5863), 588. doi: 10.1126/science.1149683

Banks, S. C., & Peakall, R. S. (2012). Genetic spatial autocorrelation can readily detect sex-biased dispersal. *Molecular Ecology, 21*(9), 2092-2105. doi: 10.1111/j.1365-294X.2012.05485.x

Boas, F. (1896). The limitations of the comparative method of anthropology. *Science, 4*(103), 901-908. doi: 10.1126/science.4.103.901

Borgerhoff Mulder, M., Nunn, C. L., & Towner, M. C. (2006). Cultural macroevolution and the transmission of traits. *Evolutionary Anthropology, 15*(2), 52-64. doi: 10.1002/evan.20088

Bottigheimer, R. B. (2009). *Fairy tales: a new history*. Albany, NY: State University of New York.

Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., . . . Atkinson, Q. D. (2012). Mapping the origins and expansion of the Indo-European language family. *Science, 337*(6097), 957-960. doi: 10.1126/science.1219669

Boyd, B. (2009). *On the origin of stories: evolution, cognition, and fiction*. London, UK: Harvard University Press.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences USA, 108*, 10918-10925. doi: 10.1073/Pnas.1100290108

Bromhan, L., Hua, X., Fitzpatrick, T. G., & Greenhill, S. J. (2015). Rate of language evolution is affected by population size. *Proceedings of the National Academy of Sciences USA, 112*(7), 2097-2102. doi: 10.1073/pnas.1419704112

Brown, D. E. (1991). *Human universals*. New York, NY: McGraw-Hill.

Brown, S., Savage, P. E., Ko, A. M., Stoneking, M., Ko, Y. C., Loo, J. H., & Trejaut, J. A. (2014). Correlations in the population structure of music, genes and language. *Proceedings of the Royal Society B-Biological Sciences, 281*(1774), 20132072. doi: 10.1098/rspb.2013.2072

Bryant, D., Filimon, F., & Gray, R. D. (2005). Untangling our past: languages, trees, splits and networks. In R. Mace, C. J. Holden, & S. Shennan (Eds.), *The evolution of cultural diversity: phylogenetic approaches*. London, UK: UCL Press.

Bryant, D., & Moulton, V. (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution, 21*(2), 255-265. doi: 10.1093/molbev/msh018

Burch Jr., E. S. (2005). *Alliance and conflict: the world system of the Inupiaq Eskimos*. Lincoln, NE: University of Nebraska Press.

Carroll, J. (2006). The human revolution and the adaptive function of literature. *Philosophy and Literature, 30*(1), 33-49. doi: 10.1353/phl.2006.0005

Coe, K., Aiken, N. E., & Palmer, C. T. (2006). Once upon a time: ancestors and the evolutionary significance of stories. *Anthropological Forum, 16*(1), 21-40. doi: 10.1080/00664670600572421

Collard, M., Buchanan, B., & O'Brien, M. J. (2013). Population size as an explanation for patterns in the Paleolithic archaeological record: more caution is needed. *Current Anthropology, 54*(S8), S388-S396. doi: 10.1086/673881

Collard, M., Ruttle, A., Buchanan, B., & O'Brien, M. J. (2013). Population size and cultural evolution in nonindustrial food-producing societies. *PLOS ONE, 8*(9), e72628. doi: 10.1371/journal.pone.0072628

Currie, T. E., Greenhill, S. J., Gray, R. D., Hasegawa, T., & Mace, R. (2010). Rise and fall of political complexity in island south-east Asia and the Pacific. *Nature, 467*(7317), 801-804. doi: 10.1038/nature09461

Davies, S. (2012). *The artful species*. Oxford, UK: Oxford University Press.

de Blécourt, W. (2012). *Tales of magic, tales in print: on the genealogy of fairy tales and the Brothers Grimm*. Manchester, UK: Manchester University Press.

Dorais, L.-J. (2010). *The language of the Inuit: syntax, semantics, and society in the Arctic*. Montreal, Canada: McGill-Queen's University Press.

Dorson, R. M. (1972). Concepts of folklore and folklife studies. In R. M. Dorson (Ed.), *Folklore and folklife: an introduction* (pp. 1-50). Chicago, Il: University of Chicago Press.

Dundes, A. (1986). The anthropologist and the comparative method in folklore. *Journal of Folklore Research, 23*(2/3), 125-146.

Eisenstein, E. L. (1979). *The printing press as an agent of change*. Cambridge, UK: Cambridge University Press.

ESRI. (2011). ArcGIS Desktop: Release 10.2. Redlands, CA: Environmental Systems Research Institute. .

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources, 10*(3), 564-567. doi: 10.1111/j.1755-0998.2010.02847.x

Fortescue, M. (1998). *Language relations across the Bering Strait: reappraising the archaeological and linguistic evidence*. London, UK: Cassell.

Fortescue, M. (2013). North America: Eskimo-Aleut linguistic history. In I. Ness (Ed.), *The encyclopedia of global human migration*. Chicester, UK: Wiley-Blackwell.

Fortescue, M., Jacobsen, S., & Kaplin, L. (2011). *Comparative Eskimo dictionary with Aleut cognates*. Chicago, Il: University of Chicago Press.

Friesen, T. M. (2013). North America: Paleoeskimo and Inuit archaeology. In I. Ness (Ed.), *The encyclopedia of global human migration*. Chicester, UK: Wiley-Blackwell.

Godfrey-Smith, P. (2012). Darwinism and cultural change. *Philosophical Transactions of the Royal Society B-Biological Sciences, 367*(1599), 2160-2170. doi: 10.1098/rstb.2012.0118

Goldberg, C. (1984). The historic-geographic method: past and future. *Journal of Folklore Research, 21*(1), 1-18.

Goody, J. R., & Watt, I. (1963). The consequences of literacy. *Comparative studies in society and history, 5*, 304-345. doi: 10.1017/s0010417500001730

Gottschall, J. (2012). *The storytelling animal: how stories make us human*. Boston, MA: Houghton Mifflin Harcourt.

Gray, R. D., Bryant, D., & Greenhill, S. J. (2010). On the shape and fabric of human history. *Philosophical Transactions of the Royal Society B-Biological Sciences, 365*(1559), 3923-3933. doi: 10.1098/rstb.2010.0162

Gray, R. D., Drummond, A. J., & Greenhill, S. J. (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science, 323*(5913), 479-483. doi: 10.1126/science.1166858

Gray, R. D., Greenhill, S., & Ross, R. M. (2007). The pleasures and perils of darwinizing culture (using phylogenies). *Biological Theory, 2*(4), 360-375. doi: 10.1162/biot.2007.2.4.360

Henrich, J. (2004). Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses: the Tasmanian case. *American Antiquity, 69*(2), 197-214. doi: 10.2307/4128416

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution, 23*(2), 254-267. doi: 10.1093/molbev/msj030

Jordan, F. M., Gray, R. D., Greenhill, S. J., & Mace, R. (2009). Matrilocal residence is ancestral in Austronesian societies. *Proceedings of the Royal Society B-Biological Sciences, 276*(1664), 1957-1964. doi: 10.1098/rspb.2009.0088

Jordan, P., & Shennan, S. (2003). Cultural transmission, language, and basketry traditions amongst the California Indians. *Journal of Anthropological Archaeology, 22*(1), 42-74. doi: 10.1016/s0278-4165(03)00004-7

Jordan, P., & Shennan, S. (2009). Diversity in hunter–gatherer technological traditions: Mapping trajectories of cultural 'descent with modification' in northeast California. *Journal of Anthropological Archaeology, 28*(3), 342-365. doi: 10.1016/j.jaa.2009.05.004

Kline, M. A., & Boyd, R. (2010). Population size predicts technological complexity in Oceania. *Proceedings of the Royal Society B-Biological Sciences, 277*(1693), 2559-2564. doi: 10.1098/rspb.2010.0452

Krohn, K. (1926/1971). *Folklore methodology [Die folkloristische Arbeitsmethode]*. Austin, TX: The University of Texas Press.

Larsen, A. W. (2011). Evolution of Polynesian bark cloth and factors influencing cultural change. *Journal of Anthropological Archaeology, 30*(2), 116-134. doi: 10.1016/j.jaa.2010.12.004

Lee, S., & Hasegawa, T. (2011). Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages. *Proceedings of the Royal Society B-Biological Sciences, 278*(1725), 3662-3669. doi: 10.1098/rspb.2011.0518

Levinson, S. C., & Gray, R. D. (2012). Tools from evolutionary biology shed new light on the diversification of languages. *Trends in Cognitive Sciences, 16*(3), 167-173. doi: 10.1016/j.tics.2012.01.007

Lewis, M. P. (2013). *Ethnologue: languages of the world* (17th ed.). Dallas, TX: SIL International.

Mace, R., & Jordan, F. M. (2011). Macro-evolutionary studies of cultural diversity: a review of empirical studies of cultural transmission and cultural adaptation. *Philosophical Transactions of the Royal Society of London B-Biological Sciences, 366*(1563), 402-411. doi: 10.1098/rstb.2010.0238

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research, 27*(2), 209-220.

Mesoudi, A. (2011). *Cultural evolution: how Darwinian theory can explain human culture and synthesize the social sciences*. Chicago, Il: University of Chicago Press.

Mullins, D. A., Whitehouse, H., & Atkinson, Q. D. (2013). The role of writing and recordkeeping in the cultural evolution of human cooperation. *Journal of Economic Behavior & Organization, 90S*, S141-S151. doi: 10.1016/j.jebo.2012.12.017

Nettle, D. (2009). Ecological influences on human behavioural diversity: a review of recent findings. *Trends in Ecology & Evolution, 24*(11), 618-624. doi: 10.1016/j.tree.2009.05.013

Ong, W. J. (1982). *Orality and literacy: the technologizing of the word*. London, UK: Routledge.

Opie, C., Shultz, S., Atkinson, Q. D., Currie, T., & Mace, R. (2014). Phylogenetic reconstruction of Bantu kinship challenges main sequence theory of human social evolution. *Proceedings of the National Academy of Sciences USA, 111*(49), 17414-17419. doi: 10.1073/pnas.1415744111

Peakall, R., Ruibal, M., & Lindenmayer, D. B. (2003). Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, Rattus fuscipes. *Evolution, 57*(5), 1182-1195. doi: 10.1111/J.0014-3820.2003.Tb00327.X

Peakall, R., & Smouse, P. E. (2006). GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes, 6*(1), 288-295. doi: 10.1111/J.1471-8286.2005.01155.X

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics, 28*(19), 2537-2539. doi: 10.1093/bioinformatics/bts460

Poe, M. T. (2010). *A history of communications: media and society from the evolutoin of speech to the Internet*. Cambridge, UK: Cambridge University Press.

Powell, A., Shennan, S., & Thomas, M. G. (2009). Late Pleistocene demography and the appearance of modern human behavior. *Science, 324*(5932), 1298-1301. doi: 10.1126/science.1170165

Pratt, K. L. (2012). Reconstructing 19th-century Eskimo-Athabascan boundaries in Unalakleet river drainage. *Arctic Anthropology, 49*(2), 94-112. doi: 10.1353/arc.2012.0031

Raghavan, M. (2014). The genetic prehistory of the New World Arctic. *Science, 345*(6200), 1255832. doi: 10.1126/science.1255832

Read, D. (2006). Tasmanian knowledge and skill: maladaptive imitation or adequate technology? *American Antiquity, 71*(1), 164-184. doi: 10.2307/40035327

Read, D. (2012). Population size does not predict artefact complexity: analysis of data from Tasmania, Arctic hunter-gatherers, and Oceania fishing groups. *UC Los Angeles: Hum. Complex Systems* http://www.escholarship.org/uc/item/61n4303q

Rogers, D. S., & Ehrlich, P. R. (2008). Natural selection and cultural rates of change. *Proceedings of the National Academy of Sciences USA, 105*(9), 3416-3420. doi: 10.1073/pnas.0711802105

Ross, R. M., & Atkinson, Q. D. (in press). Cultural differentiation does not entail group-level structure: the case for geographically explicit analysis. *Behavioral and Brain Sciences*.

Ross, R. M., Greenhill, S. J., & Atkinson, Q. D. (2013). Population structure and cultural geography of a folktale in Europe. *Proceedings of the Royal Society B-Biological Sciences, 280*(1756), 20123065. doi: 10.1098/rspb.2012.3065

Rubin, D. C. (1995). *Memory in oral traditions: the cognitive psychology of epic, ballads, and counting-out rhymes*. Oxford, UK: Oxford University Press.

Saslis-Lagoudakis, C. H., Hawkins, J. A., Greenhill, S. J., Pendry, C. A., Watson, M. F., Tuladhar-Douglas, W., . . . Savolainen, V. (2014). The evolution of traditional knowledge: environment shapes medicinal plant use in Nepal. *Proceedings of the Royal Society B-Biological Sciences, 281*(1780), 20132768. doi: 10.1098/rspb.2013.2768

Scalise Sugiyama, M. (2001). Food, foragers, and folklore: the role of narrative in human subsistence. *Evolution and Human Behavior, 22*, 221-240. doi: 10.1016/s1090-5138(01)00063-0

Sheppard, W. L. (1998). Population movements, interaction, and legendary geography. *Arctic Anthropology, 35*(2), 147-165.

Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Multiple-regression and correlation extensions of the Mantel Test of matrix correspondence. *Systematic Zoology, 35*(4), 627-632. doi: 10.2307/2413122

Sterelny, K. (2012). *The evolved apprentice*. Cambridge, MA: MIT Press.

Tehrani, J. J. (2013). The phylogeny of Little Red Riding Hood. *PLOS ONE, 8*(11), e78871. doi: 10.1371/journal.pone.0078871

Tehrani, J. J., & Collard, M. (2002). Investigating cultural evolution through biological phylogenetic analysis of Turkmen textiles. *Journal of Anthropological Archaeology, 21*, 443-463. doi: 10.1016/s0278-4165(02)00002-8

Teverson, A. (2013). *Fairy tale*. Abingdon, UK: Routledge.

Thompson, S. (1946). *The folktale*. New York, NY: Holt, Rinehart and Winston, Inc.

Towner, M. C., Grote, M. N., Venti, J., & Borgerhoff Mulder, M. (2012). Cultural macroevolution on neighbor graphs: vertical and horizontal transmission among western North American Indian societies. *Human Nature, 23*(3), 283-305. doi: 10.1007/s12110-012-9142-z

Uther, H. J. (1997). Indexing folktales: a critical survey. *Journal of Folklore Research, 34*(3), 209-220.

Uther, H. J. (2004). *The types of international folktales: a classification and bibliography, based on the system of Antti Aarne and Stith Thompson*. Helsinki, Finland: Suomalainen Tiedeakatemia, Academia Scientiarum Fennica.

Wiessner, P. W. (2014). Embers of society: firelight talk among the Ju/'hoansi bushmen. *Proceedings of the National Academy of Sciences USA, 111*(39), 14027-14035. doi: 10.1073/pnas.1404212111

Zipes, J. (2006). *Why fairytales stick: the evolution and relevance of a genre*. New York, NY: Routledge.

Zipes, J. (2012). *Irresistible fairy tale: the cultural and social history of a genre*. Princeton, NJ: Princeton University Press.

Chapter Eight: Discussion and conclusion

8.1 Discussion and conclusion

> I doubt if a single individual could be found from the whole of mankind who is wise every hour of his life and doesn't suffer from some form of insanity. The only difference is one of degree. A man who sees a gourd and takes it for a woman is called insane because this happens to very few people. (Erasmus, 1511/1993)

Are delusions only found in the "insane", or do us "normal" folk have delusions too? Are all unfounded beliefs delusions, or only those unfounded beliefs that most people think are bizarre? Long before scientists appeared on the scene, such questions have had a firm hold on inquisitive minds. In recent years scientists have started to make important contributions to this ongoing conversation. However, progress in developing scientific accounts of belief and culture has been slow, and there is still considerable debate about how to study the relationship between belief, delusion, rationality, culture, and other cognitive and historical processes in a scientifically informed fashion. Nevertheless, ongoing research is making incremental, but important, contributions to our understanding of these topics (Boyd & Richerson, 2005; Connors & Halligan, 2015; Galbraith, 2015; Mesoudi, 2011). In this thesis I have examined belief and culture from two perspectives. Firstly, I have examined belief from the perspective of cognitive psychology. In particular, I have investigated hypotheses that propose that deficits in rationality and learning are involved in the formation and maintenance of delusional beliefs. Secondly, I have examined culture from an evolutionary perspective. In particular, I have applied methods and models from population genetics and dual-inheritance theory to map out parameters of normal cultural transmission. In this final chapter, I situate my findings within the context of recent scholarship, summarise the strengths and limitations of my research, and make some suggestions for future research.

8.2 Section one: delusional belief and rationality

It is widely believed that research using the "beads task" paradigm has provided strong evidence that delusional and delusion-prone individuals tend to manifest a "jumping to conclusions" (JTC) bias. In Chapter 2, I outlined some reasons to question the strength of the evidence. In particular, I noted that the most comprehensive evaluation of the beads task literature was a narrative review that did not synthesise evidence statistically, estimate effect sizes, or test for publication bias (Garety & Freeman, 2013). In order to overcome shortcomings of earlier reviews of this literature, I conducted a meta-analysis of studies that explored the association between the "draws to decision" measure of data gathering and scores on the Peters *et al*. Delusions Inventory (PDI) measure of delusion-like belief. This meta-analysis provided evidence that there is, in fact, an association between data gathering and PDI scores, and that this association is unlikely to be due to publication bias alone. However, the overall size of the association was found to be very small. In fact, the association was so small that the vast majority of samples failed to show a statistically significant association when examined independently (i.e. 32 of 38 samples had confidence intervals that encompass zero; see Figure 2.1). These findings make an important contribution to the literature because they run contrary to widespread claims that the JTC bias is robust, and suggest that future research is needed to interrogate more carefully the conditions under which a JTC bias might manifest most strongly.

Early in Chapter 3, I highlighted a connection between research from the delusions literature and research from the reasoning literature. In the delusions literature, data gathering in the beads task is typically interpreted as indexing the degree to which individuals exhibit a jumping to conclusions bias (Garety & Freeman, 2013). And in the reasoning literature, dual-process theories of reasoning posit that Type 1, or "intuitive," processing acts like a "machine for jump to conclusions" (Kahneman, 2011, p. 86). This shared focus on jumping to conclusions suggests that beads task performance might be associated with reasoning style. In particular, I hypothesised that people who lack "analytic cognitive style"—and therefore rely heavily on Type 1 processing—gather less data in the beads task. In what I believe is the largest beads task study to date (n = 558), I investigated the extent to which analytic cognitive style and PDI independently predict data gathering in the beads task. I found that analytic cognitive style predicted data gathering independently of PDI (and independently of some other important dimensions of individual variation), but PDI did not predict data gathering independently of analytic cognitive style. Not only did PDI fail to predict data gathering when entered into a hierarchical regression (see Table 3.3), but PDI did not even show an association with data gathering in simple correlational analysis (see Table 3.2).

The failure to find evidence for an association between data gathering and PDI appears to be at odds with the results of the meta-analysis presented in Chapter 2 that did provide evidence for an association. There are a number of possible reasons for this. One is sample size—the association supported by the meta-analysis is very small, so even a sample of 558 participants might not be sufficient to reliably identify a true association of such a small magnitude. Consequently, future research should carefully consider sample size when examining the relationship between data gathering and delusion-like belief, particularly when effect sizes are expected to be small (Cumming, 2012). Choosing appropriate sample sizes for studies using the beads task and the PDI is a challenge because distributions of PDI scores and draws to decision tend to be highly skewed and include outliers, making conventional power analyses inappropriate. Another possible reason for the inconsistency is that the study reported in Chapter 3 used Mechanical Turk to recruit participants. As discussed in Chapter 3, it is possible that participants recruited using Mechanical Turk might be in a hurry to complete the study and, as a consequence, fail to pay close attention to the task, which could obscure the associations of interest. Future research could examine whether patterns of responses from Mechanical Turk studies differ from those of conventional laboratory based studies.

It has recently been suggested that positing that people with delusions manifest a JTC bias is not informative because such a claim merely redescribes the finding that people with delusions tend to request fewer beads (Corlett & Fletcher, 2014). I have some sympathy for this critique. I think that the JTC bias needs to be situated within a cognitive model of normal data gathering. By explaining data gathering in terms analytic cognitive style, I think that the beads task study presented in Chapter 3 has provided an important first step. Future research could build on this finding by examining the relationship between clinical delusions and data gathering within the context of dual-process theories of reasoning. Such a research programme would be consistent with recent research from the delusions literature (Aimola Davies & Davies, 2009; Freeman & Garety, 2014; Speechley, Woodward, & Ngan, 2013).

The studies presented in Chapter 2 and Chapter 3 share two core features that could be considered to be limitations. First, they both used the PDI to index delusion-like belief. As I discussed in Chapter 1, this is of some concern because items from the PDI do not probe beliefs directly. This might explain why I found only very small associations in a meta-analysis and no associations in an empirical study. Consistent with this possibility, a narrative review of the beads task literature noted that the JTC bias seems to be more robust when comparing groups with clinical delusions to control groups than when comparing high PDI groups to low PDI groups from the general population (Garety & Freeman, 2013). Given that the PDI is widely used in delusions research, I think that it is important that future research examines this issue carefully. Fortunately, relevant research is on the horizon. A protocol for a meta-analysis of the beads task has recently been preregistered (Taylor, Hutton, & Dudley, 2014). This meta-analysis will compare data gathering in groups with clinical delusions to control groups. Once the findings of this meta-analysis have been published we will be in a much better position to compare data gathering in clinical delusions versus controls (i.e. the upcoming meta-analysis) to data gathering across the spectrum of PDI scores (i.e. the meta-analysis reported in Chapter 3). Another direction for research that I think would be productive is the development and testing of questionaries that probe beliefs more directly than the PDI does. One such questionnaire, which has not been used in many studies yet, is the Cardiff Beliefs Questionnaire (CBQ; Pechey & Halligan, 2011). As the developers of this questionnaire note "unlike other non-clinical measures (PDI), which use a variety of terms (e.g. 'feel', 'think' or 'worry') interchangeably in their questions, all relevant CBQ questions use the term 'belief' to avoid ambiguity" (p. 108).

The second feature of the studies reported in Chapter 2 and Chapter 3 that could be a limitation is that they both used the beads task to measure data gathering. In Chapter 2, I briefly outlined some problems with the beads task. Here I would like to discuss one of these problems in greater detail because I think it is particularly important. According to the standard interpretation of the beads task, if someone asks to see few beads then they are *jumping* to conclusions on the basis of *insufficient* evidence. Approximately half of the studies included in my meta-analysis used an 85:15 ratio of beads (the so-called "easy" version of beads task). However, at this ratio, the posterior probability favouring one jar over the other becomes very high very quickly (Maher, 2004). For example, calculations using Bayes theorem reveal that if the first two beads are of the same colour then the posterior probability of one of the jars is already greater than 0.97. Consequently, in many cases it is not unreasonable for a participant to decide that they are "certain" that they know which jar the beads are being drawn from after seeing a very small number of beads and terminate data gathering. The study by Huq, Garety, and Hemsley (1988) that introduced the beads task into the delusions literature briefly acknowledged this issue:

> Although the deluded sample's response on two of the measures appears more 'Bayesian', being less subject to the 'conservatism' bias than normal, it is not possible to conclude that deluded people are better reasoners. One could argue here that an abnormal bias—a tendency to draw on little evidence and be over-confident—is cancelling out a normal bias—towards conservativism in this type of (easy) probabilistic reasoning task. Conservativism of this type may, in certain circumstances be ecologically valid, serving as a useful general strategy under

conditions of uncertainty. (Huq et al., 1988, p. 810-811)

This, I suggest, is a highly speculative interpretation of data that, to me, seem to provide evidence *against* the hypothesis that people with delusions *jump* to conclusions. It is, of course, *possible* that some degree of conservativism is "ecologically valid," but I think that the onus is on advocates of this interpretation of the beads task data to provide evidence that accurate responses to a reasoning task provide evidence for a JTC *bias*[11]. There is now a very large literature of beads task studies, and many of these studies use the "easy" version. Nevertheless, I am not aware of any thorough discussion of this crucial issue. Consequently, I suggest that it is far from clear how studies that have used the "easy" version of the beads task should be interpreted. The 60:40 "hard" version of the beads task is, I suggest, more useful precisely because one of the jars achieves a high posterior probability only after a relatively large number of beads had been drawn. This is why I used the "hard" version in my study. Nevertheless, as discussed in Chapter 3, there are other problems with the beads task that are not resolved by using the "hard" version of the task. Recently, new data gathering paradigms have been developed that address a variety of problems associated with the beads task (Speechley, Whitman, & Woodward, 2010; van der Leer, Hartig, Goldmanis, & McKay, 2015; Woodward, Moritz, Cuttler, & Whitman, 2006). These paradigms hold considerable potential, and I suggest that it would be more productive for future to focus on using these paradigms to study data gathering, rather than the beads task.

In Chapter 4, I investigated another paradigm that has recently been used to study delusions. A series of neuroimaging studies adapted the "allergist task" from the causal learning literature to test the hypothesis that delusions are caused by aberrant prediction error signalling (Corlett et al., 2004; Corlett, Honey, & Fletcher, 2007; Corlett, Murray, et al.,

[11] There are serious challenges associated with defining rationality in terms of "ecological validity" (Boudry, Vlerick, & McKay, 2015). Consequently, any theory that attempts to defend conservative responding as being "ecologically valid" and Bayesian responding as being a manifestation of underlying pathology (and in some sense "irrational") would need to grapple with some very challenging issues.

2007). In contrast to the beads task, I think that the allergist task is a very carefully designed paradigm whose results can be interpreted with considerable certainty (i.e. "learning has occurred" vs. "learning has not occurred"). However, as I argued in Chapter 4, the statistical comparisons used to index behavioural evidence for learning in Corlett and colleagues' adaptation of the task are not appropriate (Griffiths, Langdon, Le Pelley, & Coltheart, 2014). As a consequence, the neuroimaging data from these studies are very difficult to interpret. For this reason, I developed a closely matched allergist paradigm to test for learning more rigorously. Despite having a much larger sample size than the neuroimaging studies I was attempting to replicate, I found no evidence for unovershadowing (a form of retrospective revaluation) in the crucial trials, which suggests that learning via unovershadowing probably did not occur in the crucial trails in the neuroimaging studies either.

That I did not find evidence for unovershadowing is an important result because these neuroimaging studies have been widely cited as providing the strongest evidence for prediction error dysfunction in delusions (Clark, 2013; Fletcher & Frith, 2009; Hohwy, 2013). Given the tremendous influence of predictive coding theories of cognition, it is crucial that further experimental research is undertaken to investigate whether delusions really are associated with prediction error dysfunction. In addition, it would be useful if future empirical research could contribute to debates about whether prediction error theories of delusions compete with, or complement, other cognitive theories of delusions, such as the "two factor theory" (Miyazono, Bortolotti, & Broome, 2015).

8.3 Section two: transmission of culture

Ever since Darwin, scholars have noted that there are striking parallels between biological evolution and cultural change (Mesoudi, 2011). However, it has only been relatively recently that a rigorous science of cultural evolution has been developed. At the vanguard of this research has been dual-inheritance theory (Richerson & Boyd, 2005), which borrows models and theory from population genetics to study cultural transmission. In the second section of this thesis, I presented research on the parameters of normal cultural transmission in folktales (a "selectively neutral trait") within and between ethnolingustic groups (henceforth "groups").

In Chapter 5, I examined the narrative content of 700 variants of a folktale across 31 European groups. This analysis provided strong support for group-level differentiation in folktales. That is to say, on average, folktales from the same group resembled each other to a greater degree than those from different groups. Intriguingly, this differentiation was found to be an order of magnitude *greater* than genetic differentiation in comparable European groups. Nevertheless, in Chapter 5, I noted that these results must be interpreted with some caution. As population geneticists have repeatedly demonstrated, group differentiation is not necessarily caused by group structure; differentiation can be (and often is) due to non-random spatial sampling of individuals across a landscape, rather than group boundaries (Handley, Manica, Goudet, & Balloux, 2007). To explore the causes of group differentiation, I examined the extent to which group affiliation and geographic distance *independently* predicted variation at the level of the individual folktale. These analyses demonstrated that both group membership and geographic distance explained variation at the individual level, but geography was the better predictor. This is an important result because it provided evidence for genuine group structure in a selectively neutral cultural trait, but also demonstrated that group differentiation must be interpreted within a spatially explicit framework.

Chapter 6 focused on a target article by Richerson *et al*. (in press) that drew evidence from diverse sources to support the theory that cultural group selection plays an important role in human evolution. In my commentary, I focused exclusively on one of the sources of evidence: evidence for between-group cultural variation. They presented estimates of group differentiation for neighbouring groups across 29 cultural inventories (see Supplement A of Richerson et al., in press) and noted that many of these estimates are relatively high compared to estimates of genetic differentiation in humans. This, they argued, suggests that there is more scope for cultural than genetic group selection. However, as we demonstrated in Chapter 5, cultural differentiation does not entail group level structure—and it is group level structure, not cultural differentiation per se, that would provide scope for cultural group section. Using our folktale research, and research by other groups on folkmusic (Brown et al., 2014) and material culture (Shennan, Crema, & Kerig, 2015), we provided empirical evidence that Richerson *et al.* (in press) need to take geography into account to test their evidence for cultural group selection more rigorously. I hope that this commentary persuades them to undertake a geographically explicit analysis of their data soon.

In Chapter 7, I presented an analysis of folktales from 18 hunter-gatherer groups spread across 6000 km of the Arctic. The European folktale dataset analysed in Chapter 6 had properties that limit its usefulness for making inferences about general patterns of cultural evolution. As I discussed in Chapter 7, these European cultures differ from the hunter-gather groups that characterised much of our recent evolutionary past in some potentially important respects. In particular, it has been argued that writing and book printing played important roles in distributing folktales in Europe (Bottigheimer, 2009; de Blécourt, 2012). Consequently, the Arctic groups studied in Chapter 7 are better suited to making inferences about cultural transmission prior to recent cultural innovations because, prior to European colonisation, they lived traditional hunter-gatherer lifestyles and did not have writing systems. I found that linguistic relatedness and geographic proximity independently predicted similarity in folktale inventories, which provides evidence for both vertical transmission down cultural lineages and horizontal transmission between groups. Interestingly, I found that geographic proximity was the better predictor. These results are important because they provide scope for explaining how relatively small groups managed to develop and retain complex cultural innovations—at topic of considerable debate (Henrich, 2004; Read, 2012)—namely, high-bandwidth social learning across group boundaries boosted the size of the interacting pool of social learners (i.e. the "effective population size").

The folktale datasets analysed in this thesis have a number of limitations. Importantly, it is not clear what the most appropriate "grain of analysis" is for folktale data. In the case of the European data, I examined variation in narrative elements in a *single folktale*, with each group contributed at least two (and typically many more than two) folktales to the dataset. By contrast, in the case of the Arctic data, I examined the presence of *different folktales* in a group, and there was no examination of variation within folktales or variation within groups. Given that there is considerable debate about how different grains of cultural data relate to each other (Godfrey-Smith, 2012), it is not clear that that the results of these two studies can be compared directly. Consequently, the comparisons I made between cultures with writing systems and cultures without writing systems are somewhat speculative. Another difficulty associated with studying these particular folktale datasets it that they are both rather "noisy". While extracting the data from secondary source materials (i.e. Roberts, 1958 and Sheppard, 1998), I encountered a number of typological errors, and I was forced to make uncertain inferences about what data code should be used in a non-trivial number of instances. Future research could re-examine the primary source materials to correct errors in the secondary data and to explore other coding schemes at different grains of analysis (e.g., within-group variation in the Arctic data). However, re-examining primary source materials would be a significant undertaking because many are unpublished and can only be accessed visiting the appropriate archives.

8.4 Culture and belief

The research presented in this thesis was motivated by my longstanding fascination in the relationship between culture and belief. Nevertheless, as the division of my thesis into two sections demonstrates, I have not brought culture and belief together directly in my empirical research; instead, I have explored them relatively independently of each other. This was done for practical reasons. The scientific study of these topics is rather new, and we know relatively little about the processes of either normal belief formation or normal cultural transmission. Consequently, I thought it prudent to work within existing paradigms, and these paradigms have tended to examine these topics independently of each other. That being said, I think that the research strategies I have employed point towards a more unified approach. For example, recent research suggests that religious beliefs (Gervais & Norenzayan, 2012; Pennycook, Cheyne, Seli, Koehler, & Fugelsang, 2012) and moral beliefs (Pennycook, Cheyne, Barr, Koehler, & Fugelsang, 2014) can be studied productively using dual-process theories of reasoning. This means that beliefs whose content are, to a significant extent, culturally transmitted can be productively studied from this perspective. Similarly, research that examines religious beliefs and moral beliefs from an evolutionary perspective is blossoming, with many insights coming from dual-inheritance theory (Norenzayan et al., in press). I suggest that the novel approaches I have used to study folktale transmission may prove to be useful for studying the transmission of cultural and religious beliefs.

It remains to be seen whether the cultural evolutionary methods used in this thesis to study folktales will prove to make a genuine contribution to the study of delusions. As discussed in Chapter 1, according to the DSM-5 a belief can only qualify as being a delusion if "the belief is not ordinarily accepted by other members of the person's culture or subculture (i.e., it is not an article of religious faith)" (p. 819). Even if this is correct (and there is considerable debate on this issue, as I have shown in Chapter 1), cultural evolutionary methods could still help explain the transmission of the content of delusions. This is because the content of delusions frequently depends on the wider cultural context, even when other people in the culture do not share a belief in the content of the delusion. For instance, Gold and Gold (2012) have made a strong case that the content of the "Truman Show" delusion (the belief that one's life is being filmed and broadcast as a reality television show) ought to be explained, at least in part, in terms of the popularity of reality television in contemporary culture. In subsequent work, they convincingly document a variety of ways in which, in their words, “culture shapes madness” (Gold and Gold, 2014). To me, this growing appreciation of the link between cultural transmission and delusions suggests that quantification of cultural transmission using evolutionary methods could help explain when belief in talking trees or disappearing penises is a manifestation of underlying pathology, and when such beliefs are just part of what it is to be human.

8.5 References

Aimola Davies, A. M., & Davies, M. (2009). Explaining pathologies of belief. In M. R. Broome & L. Bortolotti (Eds.), *Psychiatry as cognitive neuroscience: philosophical perspectives* (pp. 285-323). Oxford, UK: Oxford University Press.

Bottigheimer, R. B. (2009). *Fairy tales: a new history*. Albany, NY: State University of New York.

Boudry, M., Vlerick, M., & McKay, R. (2015). Can evolution get us off the hook? Evaluating the ecological defence of human rationality. *Conscious and Cognition, 33*, 524-535. doi: 10.1016/j.concog.2014.08.025

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. Oxford, UK: Oxford University Press.

Brown, S., Savage, P. E., Ko, A. M., Stoneking, M., Ko, Y. C., Loo, J. H., & Trejaut, J. A. (2014). Correlations in the population structure of music, genes and language. *Proceedings of the Royal Society B-Biological Sciences, 281*(1774), 20132072. doi: 10.1098/rspb.2013.2072

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(03), 181-204. doi: 10.1017/S0140525X12000477

Connors, M. H., & Halligan, P. W. (2015). A cognitive account of belief: a tentative roadmap. *Frontiers in Psychology, 5*(1588), 1-14. doi: 10.3389/fpsyg.2014.01588

Corlett, P. R., Aitken, M. R. F., Dickinson, A., Shanks, D. R., Honey, G. D., Honey, R. A. E., . . . Fletcher, P. C. (2004). Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron, 44*(5), 877-888. doi: 10.1016/j.neuron.2004.11.022

Corlett, P. R., & Fletcher, P. C. (2014). Computational psychiatry: a Rosetta Stone linking the brain to mental illness. *The Lancet Psychiatry, 1*(5), 399-402. doi: 10.1016/s2215-0366(14)70298-6

Corlett, P. R., Honey, G. D., & Fletcher, P. C. (2007). From prediction error to psychosis: ketamine as a pharmacological model of delusions. *Journal of Psychopharmacology, 21*(3), 238-252. doi: 10.1177/0269881107077716

Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R. F., Shanks, D. R., Robbins, T. W., . . . Fletcher, P. C. (2007). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain, 130*, 2387-2400. doi: 10.1093/brain/awm173

Cumming, G. (2012). *Understanding the new statistics: effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

de Blécourt, W. (2012). *Tales of magic, tales in print: on the geneology of fairy tales and the Brothers Grimm*. Manchester, UK: Manchester University Press.

Erasmus, D. (1511/1993). *Praise of folly*. London, UK: Penguin.

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience, 10*(1), 48-58. doi: 10.1038/nrn2536

Freeman, D., & Garety, P. A. (2014). Advances in understanding and treating persecutory delusions: a review. *Social Psychiatry and Psychiatric Epidemiology, 49*(8), 1179-1189. doi: 10.1007/s00127-014-0928-7

Galbraith, N. (Ed.). (2015). *Aberrant beliefs and reasoning*. New York, NY: Psychology Press.

Garety, P. A., & Freeman, D. (2013). The past and future of delusions research: from the inexplicable to the treatable. *British Journal of Psychiatry, 203*(5), 327-333. doi: 10.1192/bjp.bp.113.126953

Gervais, W. M., & Norenzayan, A. (2012). Analytic thinking promotes religious disbelief. *Science, 336*(6080), 493-496. doi: 10.1126/science.1215647

Godfrey-Smith, P. (2012). Darwinism and cultural change. *Philosophical Transactions of the Royal Society B-Biological Sciences, 367*(1599), 2160-2170. doi: 10.1098/rstb.2012.0118

Gold, J., & Gold, I. (2012). The "Truman Show" delusion: psychosis in the global village. Cognitive Neuropsychiatry, 17(6), 455-472. Doi: 10.1080/13546805.2012.666113

Gold, J., & Gold, I. (2014). Suspicious minds: how culture shapes madness. New York, NY: Freepress.

Griffiths, O., Langdon, R., Le Pelley, M. E., & Coltheart, M. (2014). Delusions and prediction error: re-examining the behavioural evidence for disrupted error signalling in delusion formation. *Cognitve Neuropsychiatry, 19*(5), 439-467. doi: 10.1080/13546805.2014.897601

Handley, L. J. L., Manica, A., Goudet, J., & Balloux, F. (2007). Going the distance: human population genetics in a clinal world. *Trends in Genetics, 23*(9), 432-439. doi: 10.1016/j.tig.2007.07.002

Henrich, J. (2004). Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses: the Tasmanian case. *American Antiquity, 69*(2), 197-214. doi: 10.2307/4128416

Hohwy, J. (2013). *The predictive mind*. Oxford, UK: Oxford University Press.

Huq, S. F., Garety, P. A., & Hemsley, D. R. (1988). Probabilistic judgments in deluded and non-deluded subjects. *Quarterly Journal of Experimental Psychology Section A, 40*(4), 801-812. doi: 10.1080/14640748808402300

Kahneman, D. (2011). *Thinking, fast and slow.* New York, NY: Farrar, Straus and Giroux.

Maher, B. A. (2004). Delusion. In H. E. Adams & P. B. Sutker (Eds.), *Comprehensive Handbook of Psychopathology* (3rd ed., pp. 309-340). New York, NY: Springer.

Mesoudi, A. (2011). *Cultural evolution: how Darwinian theory can explain human culture and synthesize the social sciences*. Chicago, Il: University of Chicago Press.

Miyazono, K., Bortolotti, L., & Broome, M. R. (2015). Prediction-error and two-factor theories of delusion formation *Aberrant beliefs and reasoning* (pp. 34-54). London, UK: Psychology Press.

Norenzayan, A., Shariff, A., Willard, A. K., Slingerland, E., Gervais, W. M., McNamara, R. A., & Henrich, J. (in press). The cultural evolution of prosocial religions. *Behavioral and Brain Sciences, Advance online publication*. doi: 10.1017/s0140525x14001356

Pechey, R., & Halligan, P. (2011). The Prevalence of Delusion-Like Beliefs Relative to Sociocultural Beliefs in the General Population. *Psychopathology, 44*(2), 106-115. doi: 10.1159/000319788

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2014). The role of analytic thinking in moral judgments and values. *Thinking & Reasoning, 20*(2), 188-214. doi: 10.1080/13546783.2013.865000

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition, 123*(3), 335-346. doi: 10.1016/j.cognition.2012.03.003

Read, D. (2012). Population size does not predict artefact complexity: analysis of data from Tasmania, Arctic hunter-gatherers, and Oceania fishing groups. *UC Los Angeles: Human Complex Systems* http://www.escholarship.org/uc/item/61n4303q

Richerson, P., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., . . . Zefferman, M. R. (in press). Cultural group selection plays an essential role in explaining human cooperation: a sketch of the evidence. *Behavioral and Brain Sciences, Advance online publication*. doi: 10.1017/s0140525x1400106x

Richerson, P., & Boyd, R. (2005). *Not by genes alone: how culture transformed human evolution*. Chicago, Il: University of Chicago Press.

Roberts, W. E. (1958). *The tale of the kind and the unkind girls: Aa-Th 480 and related tales*. Germany: De Gruyter.

Shennan, S. J., Crema, E. R., & Kerig, T. (2015). Isolation-by-distance, homophily, and "core" vs. "package" cultural evolution models in Neolithic Europe. *Evolution and Human Behavior, 36*, 103-109. doi: 10.1016/j.evolhumbehav.2014.09.006

Sheppard, W. L. (1998). Population movements, interaction, and legendary geography. *Arctic Anthropology, 35*(2), 147-165.

Speechley, W. J., Whitman, J. C., & Woodward, T. S. (2010). The contribution of hypersalience to the "jumping to conclusions" bias associated with delusions in schizophrenia. *Journal of Psychiatry & Neuroscience, 35*(1), 7-17. doi: 10.1503/jpn.090025

Speechley, W. J., Woodward, T. S., & Ngan, E. T. (2013). Failure of conflict to modulate central executive network activity associated with delusions in schizophrenia. *Frontiers in Psychiatry, 4*, e113. doi: 10.3389/fpsyt.2013.00113

Taylor, P., Hutton, P., & Dudley, R. (2014). Rationale and protocol for a systematic review and meta-analysis on reduced data gathering in people with delusions. *Systematic Reviews, 3*(1), 44. doi: 10.1186/2046-4053-3-44

van der Leer, L., Hartig, B., Goldmanis, M., & McKay, R. (2015). Delusion-proneness and 'jumping to conclusions': relative and absolute effects. *Psychological Medicine, 45*(6), 1253-1262. doi: 10.1017/s0033291714002359

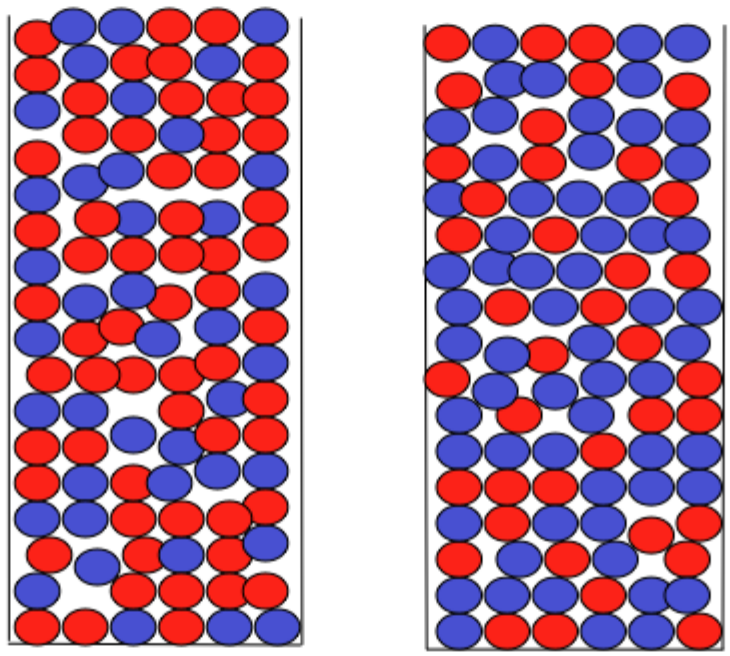
Woodward, T. S., Moritz, S., Cuttler, C., & Whitman, J. C. (2006). The contribution of a cognitive bias against disconfirmatory evidence (BADE) to delusions in schizophrenia. *Journal of Clinical and Experimental Neuropsychology, 28*(4), 605-617. doi: 10.1080/13803390590949511

Appendix: Reasoning Problems Used in Chapter Three

Beads Task

Source: Adapted from materials that were used by (Garety et al., 2005). See http://www.psychosisresearchpartnership.org.uk/tasks.html for original source materials.

Screen 1



Mainly Red Jar (60 red; 40 blue)

Mainly Blue Jar (60 blue; 40 red)

PLEASE READ THESE INSTRUCTIONS CAREFULLY:

Two jars of beads are shown above. The first jar contains 60 red beads and 40 blue beads; this is the "**mainly red jar**". The second jar contains 60 blue beads and 40 red beads; this is the "**mainly blue jar**".

I am going to hide the two jars from view. Then I am going to stir up the beads in the jars. Then I am going to draw a series of beads from ONE of these jars one at a time. Each time I draw a bead from the jar I will show you its color. Then I will return the bead to the SAME jar. Then I will stir up the beads in that jar again. Importantly, I will NEVER switch jars but will ALWAYS draw beads from the same jar.

Your task is to stop me from drawing beads as soon as you feel confident that you can guess which jar the beads are being draw from. Each time I draw a bead I will give you the option to either "decide now" or "see another bead". If you choose to "decide now" then I will stop drawing bead and will ask you to decide whether I had been drawing beads from the mainly red jar OR the mainly blue jar.

Remember a) you can see as many beads as you like before making a decision, and b) you should make your decision as soon as you feel confident that you can guess which jar the beads are being drawn from.

When you are confident that you understand these instructions and are ready to begin the task please move onto the next screen.

Screen 2

The 1st bead is blue. Would you like to decide which jar the beads are being drawn from now or would you like to see another bead?

- ❍ Decide now
- ❍ See another bead

Participants continue to be presented with screens with the next bead in the sequence until they either choose to “decide now” or have been presented with 50 beads. Then they are presented with the final screen.

Final screen

Which jar do you think the beads were being drawn from?

- ❍ The mainly red jar
- ❍ The mainly blue jar

Four Item Cognitive Reflection Test

Source: (Toplak, West, & Stanovich, 2013).
Correct responses in bold.

If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together?___ days **(4)**

Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class?___ students **(29)**

A man buys a pig for $60, sells it for $70, buys it back for $80, and sells it finally for $90. How much has he made?___ dollars **(20)**

Simon decided to invest $8,000 in the stock market one day in early 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has:
a) broke even in the stock market
b) is ahead of where he began
c) has lost money

Three Item Cognitive Reflection Test

Source: (Frederick, 2005).
Correct responses in bold.

A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? ___ cents **(5)**

If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ___ minutes **(5)**

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ___ days **(47)**

Heuristics and Biases Battery

Source: adapted from (Toplak, West, & Stanovich, 2011).
Correct responses in bold (unless otherwise stated).

1) Causal base-rate

The Smiths had decided that when it is time to replace their current car they will buy a Swedish car because they have heard that Swedish cars have a reputation for being reliable and safe. They will buy either a Volvo or a Saab.

As luck would have it, their current car broke down on the last day of a clearance sale at both the Volvo and the Saab dealerships. For this reason, if they waited another day to buy either a Volvo or a Saab, it would cost them substantially more, about $2500.

They read their Consumer Reports magazine and found that the consensus of the experts was that both cars were very sound mechanically, although the Volvo was felt to be slightly superior on some dimensions. They also read that readers of Consumer Reports who owned a Volvo reported having somewhat fewer mechanical problems than owners of Saabs.

They were about to visit the Volvo dealer to buy a car when Mr Smith remembered that they had two friends who owned a Saab and one who owned a Volvo. So Mr Smith called up these friends. Both Saab owners reported having had a few mechanical problems but nothing major. By contrast, the Volvo owner exploded when asked about his car: “First that fancy fuel injection computer stopped working and had to be replaced for 500 bucks! Next I started having trouble with the exhaust and had to replace it, and then the transmission, and then the clutch! I finally sold it after three years for junk!”

Given that the Smiths are going to buy either a Volvo or a Saab today to save about $2500, which do you think they should buy?
a) They should definitely buy the Saab.
b) They should probably buy the Saab.
c) They should probably buy the Volvo.
d) They should definitely buy the Volvo.

2) Sample size: Hospital problem

A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50 percent of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50 percent, sometimes lower. For a period of 1 year, each hospital recorded the days on which more than 60 percent of the babies born were boys. Which hospital do you think recorded more such days?
a) The larger hospital.
b) The smaller hospital.
c) About the same (that is, within 5% of each other).

3) Sample size: Squash problem

A game of squash can be played either to 9 or to 15 points (that is, the game can either end when one of the two squash players scores 9 points, or it can end when one of them scores 15 points). Holding all other rules of the game constant, if A is a better player than B, which scoring system will give A a better chance of winning?'
a) 9 points.
b) 15 points.

4) Regression to the mean

After the first 2 weeks of the major league baseball season, newspapers begin to print the top 10 batting averages. Typically, after 2 weeks, the leading batter often has an average of about .450. However, no batter in major league history has ever averaged .450 at the end of the season. Why do you think this is? Choose one:
a) When a batter is known to be hitting for a high average, pitchers bear down more when they pitch to him.
b) Pitchers tend to get better over the course of a season, as they get more in shape. As pitchers improve, they are more likely to strike out batters, so batters' averages go down.
c) A player's high average at the beginning of the season may be just luck. The longer season provides a more realistic test of a batter's skill.
d) A batter who has such a hot streak at the beginning of the season is under a lot of stress to maintain his performance record. Such stress adversely affects his playing.
e) When a batter is known to be hitting for a high average, he stops getting good pitches to hit. Instead, pitchers "play the corners" of the plate because they don't mind walking him.

5) Gambler's fallacy 1

When playing slot machines, people win approximately 1 in every 10 times. Julie, however, has just won on her first three plays. What are her chances (out of 10) of winning the next time she plays?___ out of 10

Answer:

The correct response, 1 out of 10, was scored as correct, and all other responses were scored as incorrect.

6) Gambler's fallacy 2

Imagine that we are tossing a fair coin (a coin that has a 50/50 chance of coming up heads or tails) and it has just come up heads 5 times in a row. For the 6th toss do you think that:
a) It is more likely that tails will come up than heads.
b) It is more likely that heads will come up than tails.
c) Heads and tails are equally probable on the sixth toss.

7) Conjunction problem

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in "Occupy Wall Street" demonstrations.

Please rank all the following possibilities in terms of likelihood from 1st to 8th. (That is, rank them by pulling the most likely options to the top and pulling the least likely options to the bottom)
Linda is a teacher in elementary school.
Linda works at a bookstore and takes Yoga classes.
Linda is an active feminist.
Linda is a psychiatric social worker.
Linda is a member of the League of Women Voters.
Linda is a bank teller.
Linda is an insurance salesperson.
Linda is a bank teller and is an active feminist.

Answer: Responses indicating that the conjunction was more likely than one of its components were scored as incorrect, and all other responses were scored as correct.

8) Covariation detection

A doctor had been working on a cure for a mysterious disease. Finally, he has created a drug that he thinks will cure people of the disease. Before he can begin to use it regularly, he has to test the drug. He selected 300 people who had the disease and gave them the drug to see what happened. He selected 100 people who had the disease and did not give them the drug in order to see what happened. The table below indicates what the outcome of the experiment was:

	Cure	
	Yes	**No**
Treatment present	200	100
Treatment absent	75	25

Was the treatment positively or negatively associated with the cure for this disease? Indicate your answer by choosing the appropriate number on the following scale:
−10 (strong negative association) to +10 (strong positive association)

Answer:
Negative judgments, which indicated the inefficacy of the treatment, were scored as correct.

9) Methodological reasoning

The city of Middleopolis has had an unpopular police chief for a year and a half. Many people suspect that he got appointed because he is a friend of the mayor, and he had little previous experience in police administration when he was appointed.

The mayor has recently defended this police chief in public, announcing that since he took office crime rates have decreased by 12%.

Which of the following pieces of evidence would most deflate the mayor's claim that his chief is competent?

a) The crime rates of the two cities closest to Middleopolis in location and size have decreased by 18% in the same period.
b) An independent survey of the citizens of Middleopolis shows that 40% more crime is reported by respondents in the survey than is reported in police records.
c) Common sense indicates that there is little a police chief can do to lower crime rates. The changes are probably due to social and economic conditions beyond the control of officials.
d) The police chief has been discovered to have business contacts with people who are known to be involved in organized crime.

10) Bayesian reasoning

Step 1: Imagine you are going to meet David Maxwell. Your task is to assess the probability that he is a university professor based on some information that you will be given. This will be done in two steps. At each step you will get some information that you may or may not find useful in your assessment. You are told that David Maxwell attended a party in which 25 male university professors and 75 male business executives took part, 100 people all together.

What do you think the probability is that David Maxwell is a university professor? (0-100)

Step 2: Next, you are told that David Maxwell is a member of the Bear's Club. 70% of the male university professors at the above mentioned party were members of the Bear's Club, and 90% of the male business executives at the party were members of the Bear's Club.

What do you think the probability is that David Maxwell is a university professor? (0-100)

Answer: Any decrease in probability estimate for Step 2 relative to Step 1 was scored as correct. All increases were scored as incorrect.

11) Framing problem

Step 1: Imagine that the United States is preparing for the outbreak of an unusual Asian disease that is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the program are as follows. Which of the two programs do you favor?

a) If program A is adopted, 200 people will be saved.
b) If program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.

Step 2: Again, imagine that the United States is preparing for the outbreak of an unusual Asian disease that is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the program are as follows. Which of the two programs do you favor?
a) If program C is adopted, 400 people will certainly die.
b) If program D is adopted, there is a one-third probability that no one will die and a two-thirds probability that 600 people will die.

Answer: Choosing a) at step 1 and a) at step 2 or b) at step 2 and b) was scored as correct. Other choices were scored as incorrect.

12) Probabilistic reasoning: Denominator neglect

Imagine that you have been presented with two trays that each contain a mixture of black and white marbles: a large tray that contains 100 marbles and a small tray that contains 10 marbles. The marbles are spread in a single layer on each tray.

You must draw out one marble (without peeking, of course) from either tray. If you draw a black marble, you win $2. If you draw a white marble you win nothing.

Consider a case in which the small tray contains 1 black marble and 9 white marbles, and the large tray contains 8 black marbles and 92 white marbles. Which tray should you draw a marble from if you want to win the $2?
a) The small tray (10 marbles).
b) The large tray (100 marbles).

13) Probability matching

Imagine that a dice with 4 red faces and 2 green faces will be rolled 60 times. Before each roll you will be asked to predict which color (red or green) will show up. You will be given one dollar for each correct prediction. What strategy should you use in order to make as much money as possible by making the most correct predictions?

a) Strategy A: Go by intuition, switching when there has been too many of one color or the other.
b) Strategy B: Predict the more likely color (red) on most of the rolls but occasionally, after a long run of reds, predict a green.
c) Strategy C: Make predictions according to the frequency of occurrence (4 of 6 for red and 2 of 6 for green). That is, predict twice as many reds as greens.
d) Strategy D: Predict red on all of the 60 rolls.
e) Strategy E: Predict green on all of the 60 rolls.
f) Strategy F: Predict more red than green, but switching back and forth depending upon "runs" of one color or the other.

14) Sunk cost

Step 1: Imagine that you are staying in a hotel on vacation. You paid $6.95 to see a movie on pay TV. After 5 minutes, you are pretty bored with the film. Would you watch the movie or not?
a) Watch the movie.
b) Don't watch the movie.

Step 2: Again, imagine that you are staying in a hotel on another vacation. You turn on the TV and there is a movie on. After 5 minutes, you are pretty bored with the film. Would you watch the movie or not?
a) Watch the movie.
b) Don't watch the movie.

Answer:
Responses were scored as correct if they chose a) at step 1 and a) and step 2 or b) at step 1 and b) at step 2. Other choices were coded as incorrect.

15) Outcome bias

A 55-year-old man has a heart condition. He had to stop working because of chest pain. He enjoyed his work and did not want to stop. His pain also interfered with other things, such as travel and recreation. A type of bypass operation would relieve his pain and increase his life expectancy from age 65 to age 70. However, 8% of the people who have this operation die from the operation itself.

Case 1: His physician decided to go ahead with the operation. The operation succeeded. Evaluate the physician's decision to go ahead with the operation. Please evaluate this decision on the following 7-point scale:
a) Clearly correct, and the opposite decision would be inexcusable.
b) Correct, all things considered.
c) Correct, but the opposite would be reasonable too.
d) The decision and its opposite are equally good.
e) Incorrect, but not unreasonable.
f) Incorrect, all things considered.
g) Clearly incorrect, and the opposite decision would be inexcusable.

Case 2: His physician decided to go ahead with the operation. The operation failed and the man died. Evaluate the physician's decision to go ahead with the operation. Please evaluate this decision on the following 7-point scale:
a) Clearly correct, and the opposite decision would be inexcusable.
b) Correct, all things considered.
c) Correct, but the opposite would be reasonable too.
d) The decision and its opposite are equally good.
e) Incorrect, but not unreasonable.
f) Incorrect, all things considered.
g) Clearly incorrect, and the opposite decision would be inexcusable.

If the evaluation of the decision for Case 1 is the same as for Case 2 then the response is coded as correct. Otherwise it is coded as incorrect.

Syllogisms

Source: (De Neys & Franssens, 2009).
Correct responses in bold.

Premise 1: All things that are smoked are bad for your health
Premise 2: Cigarettes are smoked
Conclusion: Therefore, cigarettes are bad for your health
Assume that the two premises are true. Does the conclusion logically follow from the two premises?
a) Yes
b) No

Premise 1: All vehicles have wheels
Premise 2: A boat is a vehicle
Conclusion: Therefore, a boat has wheels
Assume that the two premises are true. Does the conclusion logically follow from the two premises?
a) Yes
b) No

Premise 1: All flowers need water
Premise 2: Roses need water
Conclusion: Therefore, roses are flowers
Assume that the two premises are true. Does the conclusion logically follow from the two premises?
a) Yes
b) No

Premise 1: All birds have wings
Premise 2: Crows are birds
Conclusion: Therefore, crows have wings
Assume that the two premises are true. Does the conclusion logically follow from the two premises?
a) Yes
b) No

Premise 1: All African countries are warm
Premise 2: Spain is warm
Conclusion: Therefore, Spain is an African country
Assume that the two premises are true. Does the conclusion logically follow from the two premises?
a) Yes
b) No

Premise 1: All things with an engine need oil
Premise 2: Cars need oil
Conclusion: Therefore, cars have engines
Assume that the two premises are true. Does the conclusion logically follow from the two premises?
a) Yes
b) No

Premise 1: All meat products can be eaten
Premise 2: Apples can be eaten
Conclusion: Therefore, apples are meat products
Assume that the two premises are true. Does the conclusion logically follow from the two premises?
a) Yes
b) No

Premise 1: All mammals can walk
Premise 2: Whales are mammals
Conclusion: Therefore, whales can walk
Assume that the two premises are true. Does the conclusion logically follow from the two premises?
a) Yes
b) No

Wordsum

Source: (Huang & Hauser, 1998).
Correct responses in bold.

We would like to know something about how people go about guessing words they do not know. Below you can see ten words in capital letters, like BEAST. You may know some of them, and you may not know quite a few of them. There are five words underneath each of these words. Please choose the word that comes closest to the meaning of the word in capital letters.

For example, if the word in capital letters is BEAST, you would choose "animal" because it comes closer in meaning to BEAST than any of the other words.

BEAST 1. afraid 2. words 3. large 4. animal 5. separate

A. SPACE 1. school 2. noon 3. captain **4. room** 5. board
B. BROADEN 1. efface 2. make level 3. elapse 4. embroider **5. widen**
C. EMANATE 1. populate 2. free 3. prominent 4. rival **5. come**
D. EDIBLE 1. auspicious 2. eligible **3. fit to eat** 4. sagacious 5. able to speak
E. ANIMOSITY **1. hatred** 2. animation 3. disobedience 4. diversity 5. friendship
F. PACT 1. puissance 2. remonstrance **3. agreement** 4. skillet 5. pressure
G. CLOISTERED 1. miniature 2. bunched 3. arched 4. malady **5. secluded**
H. CAPRICE 1. value 2. a star 3. grimace **4. whim** 5. inducement
I. ACCUSTOM 1. disappoint 2. customary 3. encounter **4. get used to** 5. business
J. ALLUSION **1. reference** 2. dream 3. eulogy 4. illusion 5. aria
H. AUDACIOUS **1. Daring** 2. Smart 3. Brave 4. Loud 5. Outgoing

I. ENCUMBER **1. Impede** 2. Oppress 3. Gather 4. Press 5. Encompass

Numeracy

Source: (Schwartz, Woloshin, Black, & Welch, 1997).
Correct responses in bold.

We would like to know something about how people make numerical judgments about uncertainty. Please answer the following three problems.

Imagine that we flip a fair coin 1,000 times. What is your best guess about how many times the coin would come up heads in 1,000 flips?____ times out of 1,000 **(500)**

In the BIG BUCKS LOTTERY, the chances of winning a $10 prize is 1%. What is your best guess about how many people would win a $10 prize if 1000 people each buy a single ticket to BIG BUCKS?____ person(s) out of 1,000 **(10)**

In ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets to ACME PUBLISHING SWEEPSTAKES win a car?____% of tickets **(0.1)**

References

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition, 113*(1), 45-61. doi: 10.1016/j.cognition.2009.07.009

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25-42. doi: 10.1257/089533005775196732

Garety, P. A., Freeman, D., Jolley, S., Dunn, G., Bebbington, P. E., Fowler, D. G., . . . Dudley, R. (2005). Reasoning, Emotions, and Delusional Conviction in Psychosis. *Journal of Abnormal Psychology, 114*(3), 373-384. doi: 10.1037/0021-843X.114.3.373

Huang, M. H., & Hauser, R. M. (1998). Trends in Black-White test score differentials: II The WORDSUM vocabulary test. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 303-332). Washington, DC: American Psychological Association.

Schwartz, L. M., Woloshin, S., Black, W., & Welch, H. G. (1997). The role of numeracy in understanding the beneft of screening mammography. *Annuals of Internal Medicine, 127*(11), 966-972.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39*(7), 1275-1289. doi: 10.3758/s13421-011-0104-1

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, 20(2), 1-22. doi: 10.1080/13546783.2013.844729