

# The Role of Adviser Confidence during Advice-Taking



So Young Woo  
Pembroke College  
University of Oxford

A thesis submitted for the degree of  
*Doctor of Philosophy*  
Trinity Term 2025

For my mother, Sang Wan Park, and my father, Byong Yol Woo. This DPhil would not have been possible without their unwavering love and support.

# Acknowledgements

First and foremost, I would like to thank my supervisor, Nick Yeung. There were tough moments in my DPhil journey, but I was able to continue thanks to your guidance. Although I still have a long way to go, I believe that I have grown as researcher, and I am truly grateful that I did my PhD under your supervision. I would also like to thank members of the ACC lab, past and present, I will miss our lab meetings and journal clubs and socials filled with cakes (thank you Nick!). I thoroughly enjoyed learning about everyone's research, which also helped me think more broadly and reflect on my own research. Special thanks to Sriraj and Samuel, I would not have been able to do my DPhil without your help with coding and our trips to the pub.

I am also grateful to the people I met at Oxford. I would like to thank everyone who helped me throughout my DPhil by listening to me rant, give me advice and forcing me to take breaks. Special thanks to Yishan for all the encouragement and the laughs we shared. You are the best flatmate anyone could ask for and I hope we can live together again in the future. I would also like to thank my ISP friends, Kelly and Aayush, that I can't believe after 10 years after graduation I still spend most of my free time with (I should really branch out and make new friends). Thank you for always being on the other end of the line when I was feeling stressed.

Lastly, I would like to thank my supportive family, especially my parents Sang Wan Park and Byong Yol Woo. I literally wouldn't be here if it weren't for my mum, who told me to submit my application when I was hesitant. Thank you both for your financial and emotional support for the past decades. I feel extremely privileged to be your daughter and love you dearly.

# Abstract

People often seek advice during decision-making and people consistently exhibit preference for confidently given advice. Recent work suggests that advisees' evaluation of adviser confidence is not one-dimensional and considers a variety of factors, such as the level of uncertainty in the decision-making context. The current thesis extends this notion and examines whether advisees make complex inferences about adviser confidence in three experimental chapters. The first line of research investigates whether adviser confidence is evaluated in relation to the advice-taking context and if advisees are sensitive to the subtle cues of competence embedded in adviser confidence. More specifically, advisees' perception and advice adoption of advice given in context-appropriate linguistic expressions was investigated. Across five experiments, advisees showed limited attention to the congruency between advisers' linguistic expression and type of uncertainty prevalent in the advice-taking context. The second experimental chapter explored how the discrepancy between external information ambiguity and adviser confidence is processed. Three experiments examined the processing of highly confident advice when external information is variable and the impact of uncertain advice despite clear external information. The results showed that in the former situation, advisees tend to become less confident when an adviser disagrees with their initial answer but increase their confidence when an adviser agrees. Conversely, in the latter scenario, advisees often decreased their confidence in spite of an agreeing advice. This behaviour was most consistently observed when the advisees were considerably overconfident compared to an expert (vs. novice) adviser. Furthermore, advisees showed hints of learning and adopting the advisers' use of the external information during decision-making. The results suggest that adviser confidence is used flexibly to understand the external information and to evaluate one's own performance. Advisees' tendency to learn advisers' use of the external information from adviser confidence was not the main objective in the second experimental chapter. Therefore, the third line of work further examined this tendency through experimental set-up that resembled cue learning. It was

hypothesised that adviser confidence would be perceived as a proxy for performance feedback and help advisees to evaluate the external information. However, two experiments found limited signs of learning from adviser confidence. Overall, the experiments in the current thesis suggests that adviser confidence is not processed in isolation and advisees may be sensitive to subtle factors of adviser confidence under specific circumstances.

## **Table of Contents**

<b>Chapter 1: Introduction</b>	<b>1</b>
<b>1.1 Social Influence and Advice</b>	<b>1</b>
1.1.1 Unique Features of Advice	2
1.1.2 Definition of Advice	4
<b>1.2 Rationale for Advice-Taking</b>	<b>6</b>
1.2.1 Information Seeking	6
1.2.2 Reduce Decision Bias	6
1.2.3 Social Benefits	7
<b>1.3 Research methods in advice-taking</b>	<b>8</b>
1.3.1 Judge Adviser System	9
1.3.2 Measures of Advice-adoption	10
<b>1.4 Egocentric discounting</b>	<b>13</b>
<b>1.5 Expertise and Advice Adoption</b>	<b>16</b>
1.5.1 Dimensions of expertise	16
1.5.2 Advisees' expertise	17
1.5.3 Advisers' expertise	18
<b>1.6 Confidence and advice adoption</b>	<b>19</b>
1.6.1 Definition of confidence	19
1.6.2 Advisee confidence and advice adoption	20
1.6.3 Adviser confidence and advice adoption.	20
1.6.4 Confidence Heuristic	21
1.6.5 Confidence and accuracy calibration	22
1.6.6. Flexible use of confidence and its link to other factors.	24
<b>1.7 Thesis Outline</b>	<b>28</b>
<b>Chapter 2: Adviser Confidence and Uncertainty in Advice-Taking Context</b>	<b>32</b>
<b>2.1 Introduction</b>	<b>32</b>
2.1.1 Research Outline	37
<b>2.2 Experiment 1</b>	<b>38</b>
2.2.1 Methods	38
2.2.2 Results	46
2.2.3 Discussion	52
<b>2.3 Experiment 2</b>	<b>55</b>
2.3.1 Methods	55
2.3.2 Results	59
2.3.3 Discussion	65
<b>2.4 Experiment 3</b>	<b>67</b>
2.4.1 Methods	67
2.4.2 Results	73
2.4.3 Discussion	78
<b>2.5 Experiment 4</b>	<b>79</b>

2.5.1 Methods	80
2.5.2 Results	85
2.5.3 Discussion	91
<b>2.6 Experiment 5</b>	<b>93</b>
2.6.1 Methods	93
2.6.2 Results	95
2.6.3 Discussion	101
<b>2.7 General Discussion</b>	<b>102</b>
2.7.1 Limitations	107
2.7.2 Conclusion	109
<b>Chapter 3: External Information Ambiguity and Adviser Confidence</b>	<b>110</b>
<b>3.1 Introduction</b>	<b>110</b>
3.1.1 Research Outline	116
<b>3.2 Experiment 6</b>	<b>117</b>
3.2.1 Methods	118
3.2.2 Results	128
3.2.3 Discussion	139
<b>3.3 Experiment 7</b>	<b>141</b>
3.3.1 Methods	141
3.3.2 Results	144
3.3.3 Discussion	154
<b>3.4 Experiment 8</b>	<b>157</b>
3.4.1 Methods	157
3.4.2 Results	159
3.4.3 Discussion	168
<b>3.5 Meta-analysis of effect size</b>	<b>171</b>
<b>3.6 General Discussion</b>	<b>173</b>
3.5.1 Limitations	179
3.5.2 Conclusions	180
<b>Chapter 4: Learning External Information Validity from Adviser Confidence</b>	<b>183</b>
<b>4.1 Introduction</b>	<b>183</b>
4.1.1 Research Outline	188
<b>4.2 Experiment 9</b>	<b>189</b>
4.2.1 Methods	190
4.2.2 Results	198
4.2.3 Discussion	205
<b>4.3 Experiment 10</b>	<b>208</b>
4.3.1 Methods	208
4.3.2 Results	214
4.3.3 Discussion	221
<b>4.4 General Discussion</b>	<b>224</b>

4.4.1 Conclusion	231
<b>Chapter 5: General Discussion</b>	<b>232</b>
	<b>232</b>
<b>5.1 Summary of Experiments</b>	<b>232</b>
<b>5.2 Implications</b>	<b>238</b>
5.2.1 Adviser Characteristics and Advice Adoption	238
5.2.2 Processing advice in relation to external context	240
5.2.3 Inferences people make about adviser confidence	244
<b>5.3 Limitations</b>	<b>245</b>
<b>5.4 Future directions</b>	<b>252</b>
<b>5.5 Conclusion</b>	<b>255</b>
<b>References</b>	<b>256</b>
<b>Appendices</b>	<b>276</b>

# Chapter 1: Introduction

---

## 1.1 Social Influence and Advice

Humans are inherently social animals and are constantly influenced by others. For example, people ask for recipes from neighbours, read product reviews online, and receive news from reporters. This widespread tendency to be influenced by others has sparked various branches of research. For example, in his seminal work, Asch (1951) found that people tended to conform to a majority's answer regardless of their own belief. Conformity generally leads to a superficial change in opinion or behaviour rather than an internal change (i.e., public conformity) (Deutsch & Gérard, 1955). This tendency to conform is found regardless of gender and age and across cultures, although the degree of conformity may vary depending on numerous factors (e.g., task difficulty, size of majority) (Capuano & Chekroun, 2024). Furthermore, Moscovici (1969) argued that even minorities could induce conformity when the minority is consistent and confident. Moreover, minorities can bring a deeper internal change in opinion and behaviour (i.e., conversion) (Moscovici, 1969). Another related branch of research is compliance, where people feel urged to accept a request. This request could be explicit where people are directly asked to perform a behaviour. For example, in Freedman and Fraser (1966) foot-in-the-door study, people were asked to put up a sign endorsing safe driving. On the other hand, the request to comply may be implicit such as political campaigns (Cialdini & Goldstein, 2004).

People's susceptibility to social influence is also evident in developmental psychology. Bandura (1977) proposed that children engage in social learning, where they observe and

model other's behaviour and then imitate the behaviour, demonstrating that people can learn vicariously. Another form of social influence is advice, which is the focus of the current thesis. Advice is a specific type of social influence but share similarities with the aforementioned types of social influence. For instance, receiving advice from several advisers may lead to conforming to the advice. Advice may also resemble explicit and implicit compliance depending on if advice is directed (e.g., 'you should do this') or expressed as an opinion (e.g., 'I think this is the right course of action). Furthermore, studies in advice-taking have presented previous participants' answers as advice (e.g., Harvey & Fischer, 1997; Yaniv & Milyavsky, 2007). Viewing others' answers may invoke social learning, where people opt to imitate the previous participant's answer. However, there are also unique features of advice that differentiate it from other forms of social influence. In the current chapter, I will give a general review of the research conducted in the advice giving and taking literature. For instance, widely used research methods in advice-taking research such as task format and measures of advice adoption will be outlined. Additionally, key terms (e.g., egocentric discounting) and core findings of factors that received scholarly attention will be summarised. After the general review, I will then focus on the factors (i.e., expertise, confidence) that were of interest in the current thesis.

### **1.1.1 Unique Features of Advice**

Research in judgement and decision-making (JDM) has not always made distinctions between general social influence and advice. Milyavsky and Gvili (2024) argued that research in advice-taking has viewed giving answers of previous participants and intentionally formulated advice by another person as interchangeable. However, social influence is a wider concept that includes all information from another individual or

groups and could be verbal or non-verbal (e.g., gesture) (Seppänen et al., 2007). On the other hand, advice is a specific type of social influence that is formed deliberately as a direct recommendation (Dalal & Bonnacio, 2010; Milyavsky & Gvili, 2024; Schrah et al., 2006). Milyavsky & Gvili (2024) found that social information framed as advice (rather than another person's opinion) had a greater impact on decision-makers. This was found consistently across six studies that used various tasks, such as estimating a person's age, guessing dates of historic events and general knowledge questions. Additionally, participants were twice more likely to purchase advice than opinions. Other research has also found that people prefer and are more likely to incorporate information given as advice compared to when it was obtained by observing another person's answer (Çelen et al., 2010; Hertz et al., 2021). Furthermore, participants performed significantly better after receiving advice compared to when they simply viewed other's answers (Hertz et al., 2021).

This preference for advice over other forms of social influence stems from the different levels of trust placed on the information. For instance, Watson and Morgan (2025) manipulated the advice-taking context so that the adviser and advisee pair were either in a cooperative or competitive relationship. Unsurprisingly, advice-taking was significantly lower in a competitive situation due to decreased trust in advice (Watson & Morgan, 2025). This was a logical course of action, as advisers were more likely to give misleading information in such situations (Watson & Morgan, 2025). Furthermore, people who are inherently trusting of others are more likely to seek and adopt advice compared to people with paranoia (Hertz et al., 2021). People place greater trust in advice as it is perceived to stem from an intention to help (Goldsmith & Fitch, 1997; Hertz et al., 2021). Thus, advice adoption decreases when this helping intention is in question (Milyavsky & Gvili,

2024; Watson & Morgan, 2025). This is also in line with the three-factor model of trust, which states that perceived trustworthiness is shaped by the perceived ability, benevolence and integrity of another party (Mayer et al., 1995). Framing information as advice seems to increase the perception of benevolence, which in turn increases trust and subsequent advice adoption.

In addition, a unique social feature of advice is that people prefer advice to be specifically tailored for them (Dove et al., 2020). This preference is found in various advice-taking domains such as nutrition (Bobokhidze et al., 2024; Briazu et al., 2024), exercise regimes (Nyman & Yardley, 2009; Yardley & Nyman, 2007), quitting smoking (Borland et al., 2004; Meyer et al., 2008), and mental health (Kooiman et al., 2023). Despite the abundance of social information in the area from reliable sources, such as doctors, research institutes, and governmental organisations, behaviour change is significantly more likely with tailored advice. This demonstrates that advice is not processed in the same manner as other forms social information. The greater tendency for behaviour change with directed advice is likely due to the higher level of perceived helping intention. In fact, Milyavsky and Gvili (2024) found that simply telling participants that advice was targeted towards another person reduced the perceived helping intention and advice adoption.

### **1.1.2 Definition of Advice**

Research on advice has been conducted in various disciplines such as psychology, behavioural economics, management and organisational behaviour, and communication studies (MacGeorge & Van Swol, 2018). The definition of advice widely varies across disciplines (MacGeorge & Van Swol, 2018). For example, in communication studies,

advice is considered as recommendations about what might be thought, said, or done to manage a problem (MacGeorge et al., 2008). Some researchers argue that advice can be broadly defined and include judgements, opinion, recommendation and guidance (Bonnacio & Dalal, 2006; Kämmer et al., 2023). Dalal and Bonnacio (2010) claimed that there are five types of advice. The first type is recommendation in favour of a course of action, which is the type of advice that was generally given in the current thesis. The second type is recommendation against a course of action, and thus suggestion to exclude an option. Third type is providing information about the decision options (e.g., giving pros and cons for each option). Here, an adviser is not urging advisees to make a certain decision but simply sharing their knowledge. These three types of advice are related to the decision alternatives, but advice could also be given about *how* to decide. For example, an adviser may suggest which factors to evaluate before reaching a decision. The last type of advice is providing socio-emotional support, such as recognising the hardship associated with the decision (Dalal & Bonnacio, 2010). Advisers appear to flexibly alter the type of advice they give. For example, Lee and colleagues (2023) found that advisee's greater self-disclosure prompted others to provide increased information but not emotional support. In fact, it is pertinent that advisers adjust their advice as people also have different advice-seeking motives. Yitshaki (2024) found that mentees request advice for three reasons: to learn about the mentor's experience, to receive reflexive feedback that could guide future behaviour, and for psychological and emotional support. In the current thesis, advice-taking motives were not collected, and advice was limited to one type (i.e., provides information about the decision). However, further research examining other forms of advice is generally needed in the judgement and decision-making field (Bonnacio & Dalal, 2006). For example, whether there is a synergetic effect of providing

different types of advice (e.g., giving information and emotional support) could be investigated.

## **1.2 Rationale for Advice-Taking**

### **1.2.1 Information Seeking**

Asking others for advice before decision-making is generally a good strategy. There are various benefits of adopting advice. Firstly, advisees could gain new information. Exploring the environment oneself through trial-and-error can be costly and time consuming (Schrah et al., 2006; Yaniv & Milyavsky, 2007). Instead, receiving advice from others reduces or eliminates the need to gather information oneself. For example, people could receive medical advice from doctors to better their health or seek justice by obtaining legal advice from lawyers. Indeed, researchers have found that lack of knowledge in an area drives advice-seeking behaviour. For instance, people who report having insufficient financial knowledge are more likely to ask for financial advice (Gentile et al., 2016; Kramer, 2016). On the other hand, people with higher faith in their own ability (i.e., self-efficacy) tend to seek less advice (Sherf & Morrison, 2020). Additionally, people who perceive themselves as knowledgeable in a domain are less likely to seek advice in that domain (Zhang et al., 2022). This is sensible as someone who already has significant knowledge would have little to gain from receiving advice. These findings demonstrate that one possible motive for advice-taking is to complement one's own knowledge.

### **1.2.2 Reduce Decision Bias**

A related reason for advice-taking is to reduce decision-biases that cloud one's decisions. During decision-making people are highly susceptible to factors such as past experiences

(Cohen et al., 2008; Duncan & Shohamy, 2016), culture (Yates & De Oliveira, 2016), and cognitive biases (Yoon et al., 2021). Integrating another person's opinion can reduce the errors in decision-making that arise from individual differences. Numerous research has shown that the average of multiple people's answers is remarkably close to the true value (Galton, 1907; Johnson et al., 2001). This phenomenon was termed the "wisdom of the crowd" effect (Surowiecki, 2004). Researchers suggest that integrating advice from independent multiple sources lead to similar outcomes (Lorenz et al., 2011; Van Swol, 2009). Advisees can increase their decision accuracy by taking the average of their own prediction and the advisers (Gardner & Berry, 1995; Soll & Larrick, 2009; Yaniv & Kleinberg, 2000). For instance, Yaniv and Milyavsky (2007) found that participants' accuracy increased by 33% after integrating advice from eight different sources. Some researchers suggest that incorporating opinions of even two or three advisers can increase accuracy (Budescu & Rantilla, 2000; Yaniv & Milyavsky, 2007). Indeed, consulting just two advisers still increased decision accuracy by 27%, which was only 1% lower than people who received advice from four different sources (Yaniv & Milyavsky, 2007). The benefits of integrating advice have been found in various decision-making contexts from perceptual decision-making to more applied tasks such as economic forecasting (e.g., prediction of inflation rate) (Yaniv, 2004). For instance, CEOs' willingness to seek advice has been linked to more strategically appropriate decisions and greater organisational performance (Ma et al., 2019).

### **1.2.3 Social Benefits**

The previous two reasons are related to the information-seeking motive of advice-taking but as previously discussed, there are social aspects of advice. Hence, another possible reason for advice-taking is to reduce psychological distress. Research has shown that

receiving advice gives the illusion of cooperating with others which is emotionally rewarding (Hertz et al., 2021; Pereg et al., 2024). Additionally, receiving an agreeing advice reduces decision uncertainty regardless of its effect on one's decision accuracy (Soll et al., 2022). Furthermore, advisees seek advice to share responsibility regarding the decision outcome. In group decision-making, people downgrade negative outcomes and feel less accountable (Mynatt & Sherman, 1975). This diffusion of responsibility was also found during advice-taking, where increased advice adoption resulted in greater decrease in sense of responsibility (Cooper, 2024). On the other hand, people are more likely to attribute positive outcomes to themselves which leads to greater feeling of reward (Zein et al., 2021). Paradoxically, this sense of cooperation or receiving help may also reduce advice-taking in certain cases. Specifically, people may feel that advice threatens their autonomy and fear that they would appear less competent (Liang & Xue, 2022; Paik, 2020). Thus, people in position of power are reluctant to take advice especially when they are highly competitive (Tost et al., 2012; De Wit et al., 2017, See et al., 2011).

### **1.3 Research methods in advice-taking**

As previously discussed, various disciplines conduct research in advice-taking and there is a wide disparity in the definition of advice. Therefore, it is expectable that there are variations in research methods as well. For instance, in organisational research, personnel such as managers are interviewed and asked to recall advice-taking situations (Kämmer et al., 2023). On the other hand, in judgement and decision-making (JDM) research, studies are generally conducted in labs or online where a participant performs some sort of decision-making (e.g., estimation) task (Kämmer et al., 2023). The current thesis will mainly review and adopt research methods in the JDM field.

### **1.3.1 Judge Adviser System**

In JDM research, the focus is on measuring advice-utilisation and decision-outcome after receiving advice (Kämmer et al., 2023). Therefore, studies have used a wide array of estimation tasks, such as perceptual decision-making, estimating calories in food, and investment outcome or weather forecasting (Bailey et al., 2023). These tasks generally follow the Judge Adviser System (JAS) (Sniezek & Buckley, 1995). In JAS, the decision-maker (i.e., judge) makes an initial decision and then receives advice. After receiving advice, the decision-maker may then revise their answer. The degree of advice adoption is measured by comparing the decision-maker's initial and final decisions. Throughout this thesis the decision-maker will also be referred as judge or advisee.

The above-described order is the basic form of JAS, and several variations have been used in previous research. Firstly, the structure of the task may differ. For instance, instead of using a quantitative estimation task, the decision-maker may be given a binary decision tasks (e.g., Sniezek & Buckley, 1995; Pescetelli & Yeung, 2019) or multiple-choice questions (e.g., Gibbons et al., 2003; Sniezek & Buckley, 1995; Sniezek & Van Swol, 2001). Another variation is that the judges' pre-advice answer or opinion may not be collected. Instead, judges would be given a chance to become familiar with the task and receive advice directly (e.g., Budescu & Rantilla, 2000; Sniezek & Van Swol, 2001). In some studies, judges are given performance feedback (i.e., accuracy) for themselves and/or the adviser (e.g., Fischer & Harvey, 1999). This helps advisees gauge their own performance and the adviser's ability. There also might be incentives for accuracy, which has been found to increase advice adoption (Patt et al., 2006; Sniezek & Van Swol, 2001; Sniezek et al., 2004; Wang et al., 2024). Second, the adviser's characteristics could be altered as well. Advice may not always come from one adviser but from multiple advisers

(e.g., Budescu & Rantilla, 2000; Harvey et al., 2000; Molleman et al., 2020; Yaniv & Milyavsky, 2007). The use of multiple advisers is useful in understanding how advisees integrate diverse opinions (e.g., conflicting advice). Another feature that could vary is the judge and adviser's relationship. For example, in some studies, a pair of participants are arbitrarily assigned to be a judge and an adviser. The pair of participants may have pre-existing relationships (e.g., friends) or could be complete strangers (e.g., Van Swol & Sniezek, 2005). In addition, instead of advice being generated in real-time, researchers may prepare advice in advance. One commonly used method is to present answers from previous participants as advice (e.g., Hütter & Ache, 2016). Another method is to give experimenter-generated advice (e.g., Rebholz & Hütter, 2022). This method allows greater control over the characteristics of the advice (e.g., how often an adviser agrees with a judge's initial opinion). Additionally, the advisee may or may not be aware of the source of the advice. For example, participants could be told that advice comes from another person or an artificial agent to examine the difference in advice adoption (e.g., Gaube et al., 2023). Lastly, how the advisee interacts with the adviser may vary. Participants are sometimes given the option to solicit (or not) advice (e.g., Schrah et al., 2006), which could be free or have an attached monetary cost (e.g., Patt et al., 2006). Advice is most commonly provided in written form, but small number of studies have used advice given verbally or through extended interactions (Bonnacio & Dalal, 2006). These are some examples of the variations of JAS and many different forms could be utilised.

### **1.3.2 Measures of Advice-adoption**

Research using the JAS has largely focused on measuring the degree of advice adoption. Two widely used measures in estimation tasks are Weight of Advice (WOA) (Yaniv,

2004) and Weight of Own Estimate (WOE) (Yaniv & Kleinberger, 2000). The formula for each measure is as follows:

Weight of Advice (WOA) quantifies how far an advisee shifts towards advice in relation to their own initial answer. Specifically, advice adoption is measured as the ratio between the change in judge's estimate from pre- to post-advice and the difference between advice and judge's initial estimate:

$$WOA = \frac{|judge's\ final\ estimate - judge's\ initial\ estimate|}{|adviser's\ estimate - judge's\ initial\ estimate|}$$

Thus, WOA = 0 when a decision-maker ignores advice and sticks precisely to their initial estimate in their final decision. Conversely, WOA = 1 when a decision-maker abandons their own initial estimate and adopts the advised answer exactly in their final decision.

Weight of Own Estimate (WOE) also measures advisees' use of advice in relation to their own estimate. Specifically, the weight advisees place on their own estimate while discounting advice.

$$WOE = \frac{|adviser's\ estimate - judge's\ final\ estimate|}{|adviser's\ estimate - judge's\ initial\ estimate|}$$

The WOE values also range from 0 to 1. However, here, 0 illustrates complete adoption of advice whereas 1 indicates a total disregard of advice. It is worth noting that WOE = 1 - WOA and thus, researchers generally only report WOA. The strength of the two measures is their versatility. These measures can be used in various estimation tasks that uses JAS, from perceptual estimation to weather forecasting. Additionally, advice utilisation can be measured after each decision (Bonnacio & Dalal, 2006), which facilitates comparison of

the degree of advice adoption in different settings, such as varying adviser characteristics or advice-taking contexts.

On the other hand, one weakness is that due to how the formulae are constructed, the values are undefined when the judge's initial estimate equals advice. The two formulae are optimal when the judge's final estimate falls between judge's initial estimate and advice. Studies generally report that this is the case most (approximately 95%) of the time (Bailey et al., 2023). However, there could be trials where judges push away from advice and give a final estimate that is even further away from advice than their initial estimate, or judge's may overshoot and give a final estimate more extreme than the advice (Bonnacio & Dalal, 2006). Researchers using WOA or WOE often set cutoff values and disregard such cases (Bailey et al., 2023). However, these trials may provide insight into certain behaviours such as advisees becoming overconfident following agreeing advice. Additionally, the measures are only suitable when answers are collected on a linear continuous scale (Bonnacio & Dalal, 2006). Therefore, these measures cannot be used in studies with binary decision-making tasks or multiple-choice questions.

In the current thesis, the tasks were binary decision-making tasks, but an accompanying confidence level was also collected on a continuous scale (similar to an estimation task). For example, participants were given an anchor temperature and predicted if the future temperature in a specific city will be lower or higher than the anchor temperature. Participants' answers were collected on a scale that ranged from 100% confident lower to 100% confident higher. Therefore, these types of tasks resemble estimation tasks, but one key difference is that participants' initial answer and advice may agree or disagree. In contrast, estimation tasks only measures the distance between participants initial

answer and advice (e.g., denominator in WOA and WOE). Therefore, a different measure of advice adoption will be used: adviser influence. This measure was adopted from Pescetelli and Yeung (2021)

$$\text{Adviser influence} = M\Delta_{\text{confidence for agreement}} - M\Delta_{\text{confidence for disagreement}}$$

To calculate, participants' change in confidence is first collected in each trial. After, the mean change in confidence is calculated separately for trials where participants' initial answer and advice agreed and disagreed. The difference between the two means will represent the overall adviser influence. It is generally expected that when an adviser agrees, judges will become more confident. In contrast judges would decrease their confidence when an adviser disagrees. Therefore, adviser influence is expected to be positive overall.

In contrast to WOA and WOE, this measure eliminates cases where the value would be undefined and there is no cut off values. Therefore, advice adoption can be recorded in all trials, such as when advisees push away from advice or give answers more extreme than the adviser. Adviser influence will be used as the main measure of advice adoption in the current thesis. However, in some experiments, the WOA values will also be calculated for comparison with the adviser influence measure and studies in previous literature that used estimation tasks.

## **1.4 Ego-centric discounting**

As previously discussed, there are various benefits of adopting advice. Thus, the rational course of action is to take the average between one's own answer and the adviser's answer (Gardner & Berry, 1995; Yaniv & Kleinberg, 2000). However, advisees generally shift their answer towards the adviser's only about 20 to 40%, demonstrating people's

tendency to place greater weight on their own opinion (Bailey et al., 2023; Soll & Larrick, 2009). This phenomenon was termed *egocentric discounting* (Yaniv & Kleinberg, 2000). Egocentric discounting has been found consistently in numerous studies in judgement and decision making. Some researchers argue that egocentric discounting occurs as one's own answer acts as an anchor (Harvey & Fischer, 1997, Lim and O'Connor, 1995; Yaniv & Milyavsky, 2007). However, egocentric discounting has some differences from anchoring. The effect of an anchor is temporary and short-term whereas egocentric discounting is a long-term preference and over-weighing of own opinion (Harvey & Harries, 2004). In support of this, even when advisees are not asked to give their initial answer in a JAS task, egocentric discounting reduces but it is not eliminated (Krueger, 2003). Furthermore, Harvey & Harries (2004) found that judges placed greater weight on another person's opinion when it was labelled as their own opinion.

There are two prominent reasons for egocentric discounting that received scholarly attention. One possible reason is that people believe that their own ability or opinion is superior compared to an adviser (Krueger, 2003). In general, peoples' self-evaluation is biased where people think their abilities, attributes and personal traits are above average (i.e., better-than-average-effect) (Alicke et al., 1995; Alicke & Govorun, 2005; Zell et al., 2020). Advisees may believe that their performance is above average regardless of their familiarity with the task. As previously mentioned, one reason for seeking advice is to gain information. If advisees believe that they have greater information or ability in the advice-seeking domain, it is natural to be reluctant to take advice from another who may perform worse than themselves. In support of this notion, researchers found that advice adoption increases when the perceived expertise of the adviser increases (Bonnacio & Dalal, 2010; Sniezek et al., 2004; Jungermann & Fischer, 2014; Harvey & Fisher, 1997).

A second possible motivation for egocentric discounting is that people only have access to their own thought process and reasoning (Yaniv, 2004; Van Swol, 2011). People are aware of how they have formulated their own opinion and, consequently, can retrieve more evidence supporting their own opinion compared to the adviser's opinion. Van Swol (2011) found that when an adviser provided reasoning underlying their decision, it significantly increased advice adoption, particularly when the task was related to matter of taste. Additionally, advice adoption increased when advisees and advisers discussed the task together and reached a consensual decision (Minson et al., 2011). These findings suggest that greater communication from the adviser allows advisees to get a fuller picture of the adviser's decision-making process, which in turn reduces egocentric discounting.

Egocentric discounting leads to lower decision accuracy (Yaniv & Choshen-Hillel, 2012) and greater inclination for confirmation bias (Kappes et al., 2020; Rollwage et al., 2020). Ignoring advice also hinders future opportunities for advice as advisers resent advisees who reject their advice and are sceptical of future advice-giving (Brooks et al., 2015; Treffenstädt, 2017). Therefore, diverse research has been conducted to better understand factors that could reduce egocentric discounting. Previous literature has found that advice-adoption increases when advisees feel greater responsibility regarding the decision-outcome (De Wit et al., 2017; MacGeorge et al., 2002). Advisee's perceived similarity with the adviser also leads to greater advice adherence (Feng & MacGeorge, 2010). Advisee's valence towards the adviser and own emotions also influences degree of advice adoption. For instance, advice-adoption increases when advisees feel positive emotions towards the adviser such as gratitude (De Hooze et al., 2014) or liking (Belkin & Kong, 2018) but reduces when advisees feel anger towards the adviser (Belkin & Kong, 2018; Van Swol et al., 2018). In addition, one's private emotion such as pride or shame

also drives advice adoption (Van Swol et al., 2018). It is possible that these emotions influence one's perceived ability, which is a driver of advice adoption. Another factor with similar effect would be task difficulty, where increased task difficulty leads to greater advice-taking (Schrah et al., 2006). As previously mentioned, expertise of self and adviser also influences advice taking. People who believe they have greater expertise are less willing to seek and adopt advice (Kassirer et al., 2020; Milyavsky et al., 2017), whereas advice adoption increases when an adviser holds greater expertise (Bonnacio & Dalal, 2010). Similarly, when an advisee is highly confident, they are less likely to adopt advice (Soll & Larrick, 2009), whereas highly confident advisers are more influential (Van Swol, 2009). Many factors have been widely researched in the advice-taking literature. Among these, confidence and expertise have been consistently found to play a significant role during advice-taking. Thus, these two factors were researched in the current thesis and will be reviewed in further detail.

## **1.5 Expertise and Advice Adoption**

### **1.5.1 Dimensions of expertise**

Research on expertise has been conducted in various fields such as psychology, engineering and education. The core definition of expertise is similar across disciplines, but some definitions are more complex with greater dimensions. In essence, expertise is defined as having knowledge in a domain or a task (Herling, 2000). This knowledge must be recognised to be true and shared by others (Herling, 2000). However, there should also be some sort of exclusivity or esotericity as knowledge held by everyone disallow the distinction between a layperson and an expert (Collins, 2016). Furthermore, the superior knowledge held by the expert may be from learning but also from experiences (i.e., tacit knowledge) (Collins, 2016; Jungermann & Fischer, 2014; Shanteau & Stewart, 1992). An

expert should also be able to critically evaluate and flexibly use their knowledge (Collins, 2016). This ability to apply their knowledge should be evident during problem solving, in terms of their speed and accuracy in the decision outcome (Dukas, 2019; Herling, 2000). In addition, researchers in education argue that the capability to help others and pass on their knowledge is an important feature of an expert (Kennedy, 1987). In general, research in advice-taking in psychology focus on short-term interactions (Bonnacio & Dalal, 2006), which is also the case with experiments in this thesis. These interactions do not involve complex advice-giving such as transfer of tacit knowledge. Therefore, a simple definition of expertise will be employed: task- or domain- specific knowledge (Herling, 2000).

### **1.5.2 Advisees' expertise**

Advisee's existing knowledge in the advice-seeking domain plays a significant role in advice-adoption. Advisees are more open to advice when they believe that they have lower knowledge compared to the adviser (Kassirer et al., 2020; Milyavsky et al., 2017; Önkal et al., 2017; Zhang et al., 2022). This tendency is found in various domains such as financial advice-taking (Gentile et al., 2016; Kramer, 2016) and counselling (Peng, 2022). On the other hand, experts take a more critical stance towards advice and prefer to use their own knowledge to reach a decision (Cottam & Jha, 2025). Advisees may also use their own knowledge to evaluate the adviser. For instance, Önkal and colleagues (2017) manipulated the track record and status (financial analyst vs. taxi driver) of advisers in a stock price prediction task. The studies found that participants with lower expertise focused solely on the adviser's past performance and disregarded the adviser's status. However, participants with greater knowledge in stocks used both the adviser's past performance and status to evaluate the quality of advice. In addition, Yaniv (2004) found that advisees with greater expertise are more sensitive to the magnitude of

difference between their own answer and advice. More specifically, knowledgeable advisees exhibited greater egocentric discounting for distant advice. These difference in adviser evaluation suggests that advisees with prior knowledge appraise advice in a more complex manner and are less likely to take advice at face value (Cottam & Jha, 2025). This leads to selective advice taking where advice is only adopted in certain cases. There are benefits to this strategy, such as increased decision efficiency. For instance, Harvey and Fisher (1997) found that experienced judges were less likely to take advice overall but were more likely request advice for critical decisions compared to less experienced judges. On the other hand, selective advice-taking may hinder knowledge acquisition and exacerbate decision biases (Peng, 2022; Zhang et al., 2022).

### **1.5.3 Advisers' expertise**

Expertise of the adviser is a factor that has been consistently found to have a strong impact on advice adoption (Bonnacio & Dalal, 2006). Numerous studies showed that people prefer and place greater weight on an expert adviser's opinion compared to a novice adviser's opinion (Bonnacio & Dalal, 2010; Snizek et al., 2004; Jungermann & Fischer, 2014). On average, judges shift approximately 48% towards an expert's advice compared to a mere 32% shift to untrained advisers (Bailey et al., 2023). Meshi and colleagues (2012) found that in addition to greater advice adoption with experts, judges also report greater value of expert's advice post decision-making. Furthermore, judges are quicker in adopting and making a final decision when interacting with an expert (Meshi et al., 2012). In addition to these behavioural disparities, there were also differences in neural activities depending on adviser's expertise. In particular, there was a greater activation in the reward-sensitive brain region (ventral striatum) from discovering that expert advice will be provided rather than a novice's advice. This demonstrates that advisees feel a sense of

reward from simply interacting with an expert (Meshi et al., 2012). The preference for expert advice is likely due to the anticipation for higher quality and more accurate advice (Snizek et al., 2004; Jungermann & Fischer, 2014).

## **1.6 Confidence and advice adoption**

### **1.6.1 Definition of confidence**

Confidence is another factor in advice-taking that has garnered academic attention, where both advisee and adviser confidence influence the advice-taking dynamic. JDM research has generally defined and operationalised confidence as expectation of the extent to which a decision, opinion, or recommendation is correct (Klayman et al., 1999), which is the definition that will be adopted throughout the thesis. However, it is important to note that definition of confidence varies across and even within disciplines. To illustrate, some other definitions of confidence in JDM are ‘feeling of knowing’ (Navajas et al., 2017), ‘belief about the validity of our own thoughts’ (Meyniel et al., 2015), and belief that a choice or proposition is correct (Pouget et al., 2016). Philosophers argue that confidence should be considered in light of both reliance of a persuasive message, but also confidence in oneself (i.e., self-confidence) (Rotenstreich, 1972). Some researchers, such as statisticians, perceive confidence as a likelihood of an event and an antonym of uncertainty (Pawitan & Lee, 2021). Common features of these definitions are that confidence is a subjective feeling regarding one’s accuracy. Therefore, I adopted Klayman and colleagues (1999) definition because it incorporates the shared perception of confidence across disciplines. Although it is worth noting that these definitions may be limited to human subjects as there are difficulties translating certain terms to characterise confidence in animals (e.g., feeling, thoughts) (Kepecs & Mainen, 2012).

### **1.6.2 Advisee confidence and advice adoption**

Similar to advisee's expertise, advisee confidence significantly influences advice-taking. Pescetelli and colleagues (2021) found that judges who are confident in their answer were less likely to request advice. Confidence in initial judgement is also linked to smaller shift towards the advice (i.e., lower WOA) (Soll & Larrick, 2009; Tost et al., 2012; Wang & Du., 2018). Advisee's confidence also plays a mediating role between power and advice-taking where the tendency to dismiss advice in powerful advisees is reduced when they feel less confident (See et al., 2011). Similarly, increased task difficulty leads to lower confidence, which in turn results in greater advice adoption (Ache et al., 2020; Kassirer et al., 2020). In addition, advisees' personality traits such as self-confidence and self-efficacy impacts advice adoption. For instance, narcissists, characterised by sense of superiority and self-confidence, perceive advice as less useful. This leads to lower advice adoption regardless of their confidence in initial judgement (Kausel et al., 2015). In contrast, shy people tend to refrain from asking for advice (Liang et al., 2023). Other researchers have also found that greater self-efficacy leads to reduced feedback seeking (Poston et al., 2005; Sherf & Morrison, 2020). This tendency diminishes as adviser expertise increases, because advice-taking could satisfy the motive to better themselves (Poston et al., 2005; Wei et al., 2024). Although general level of self-confidence and self-efficacy influences one's decision-confidence (Perry, 2011), the current thesis will focus on decision-confidence because the purpose is to examine advice-taking in individual decisions rather than the general tendency at a trait level.

### **1.6.3 Adviser confidence and advice adoption.**

Previous research generally finds that confident advisers are more influential. Advice accompanied by low numerical confidence or wide confidence interval are perceived

more negatively and are adopted less (Phillips 1999; Sniezek & Buckley, 1995; Sniezek & Van Swol 2001; Van Swol, 2009; Van Swol & Sniezek, 2001; Wang et al., 2025; Yaniv, 1997). For instance, employees who express greater confidence are perceived to have made greater contribution in strategy formation (Phillips, 1999). The persuasiveness of confident advice is found regardless of the decision-making context. Van Swol (2011) found that confident advice is more influential in both decisions with objectively correct answers (e.g., intellectual task) and subjective judgements (e.g., taste forecast). Although, providing confident advice had greater impact in intellectual tasks, being the sole predictor of advice adoption (Van Swol, 2011). This substantial effect of adviser confidence suggests that it would be profitable to express high levels of confidence during advice-giving. Communicating higher confidence would help people assert their decision and gain more resources (e.g., funding) (Sniezek & Buckley, 1995). Advisers also seem to be aware of the benefits of expressing higher confidence and behave accordingly. For example, in Hertz and colleagues' (2017) studies, a participant assumed the role of an adviser and competed with another adviser to have a greater influence over the advisee's decision. The studies found that advisers systematically changed their confidence, where the advisers were more overconfident if the advisee had selected the other adviser in the previous decision. This behaviour might stem from the belief that communicating higher confidence would make them appear more competent, moral and warm (Stavarova & Evan, 2019). Additionally, advisers inflate their confidence when they have a greater motivation to persuade (Van Swol, 2009) or have a higher intention to help and there is less uncertainty attached to the decision (Haran et al., 2022).

#### **1.6.4 Confidence Heuristic**

Why are confident advisers so influential? This is because of the assumptions people hold about confident advisers. People believe that confident individuals have greater

information or knowledge and are more accurate (Price & Stone 2004; Thomas & McFadyen, 1995). Price and Stone (2004) found that people consistently prefer advisers who are overly confident compared to advisers whose accuracy and confidence are well-calibrated. The effect of the confidence heuristic is robust and found across different tasks and communication methods. Pulford and colleagues (2018) found that confident individuals are more influential in various tasks (e.g., face recognition, geometric shape categorisation) and regardless of communication method (face-to-face vs. computer based). This demonstrates that confidence heuristic is prevalent in different decision-making contexts and driven by verbal (rather than non-verbal) communication (Pulford et al., 2018).

### **1.6.5 Confidence and accuracy calibration**

On the other hand, advisees should be cautious when using confidence heuristics because confidence is not a fully reliable indicator of accuracy. Although there is a positive correlation between confidence and accuracy, people are generally overconfident (Van Swol & Snizek, 2005; Sah et al., 2013, Klayman et al., 1999; Lichtenstein et al., 1982). For instance, Van Swol and Sinezek (2005) found that advisers' mean confidence was about 10% higher than their average accuracy. In Soll and Larrick (2009)'s studies, participant dyads were only successful 50% of the time in identifying the more accurate person within the dyad but were on average 74% certain that they made the correct decision. Additionally, overconfidence was higher when the participant dyads made incorrect judgements and selected the less accurate person. This tendency for overconfidence is observed in various domains. For example, students overestimate their academic performance (Clayson, 2005), entrepreneurs are wildly optimistic about their ability and success rate (Camerer & Lovallo, 1999), and eyewitnesses are frequently

overconfident (Brewer & Wells, 2006; Tenny et al., 2007). Therefore, advisees should not be over-reliant on adviser confidence.

In fact, past research shows that people's dependence on adviser confidence diminishes when adviser accuracy information is available. For example, Tenney and colleagues (2007) showed that perceived credibility of confident witnesses decrease when their testimony was found erroneous, whereas the credibility of less confident witnesses remained intact. Additionally, applicants who displayed poor calibration between their confidence and accuracy were rated negatively by recruiters (Tenny & Spellman, 2010). Therefore, the calibration hypothesis argues that one's credibility is determined by how well one's confidence is warranted (Tenney et al., 2007; Tenney et al., 2008). Support for the calibration hypothesis is also found in advice-taking. Sah and colleagues (2013) asked participants to estimate the weight of a person in a task following judge-adviser-system (JAS). The studies found that perceived adviser credibility was determined by the degree of calibration between an adviser's confidence and their accuracy. Stanciu and Fiser (2022) also found evidence that advisees were more influenced by confident and calibrated adviser over simply confident advisers. Furthermore, uncalibrated advisers are perceived negatively, and advisees may even choose to punish such advisers (Vullioud et al., 2017).

This preference for calibration between confidence and accuracy may explain why uncertain advice is preferred sometimes. Gaertig and Simmons (2018) conducted a series of studies using tasks such as sports game outcome prediction and stock price prediction. Results showed that people were not averse to uncertain advice that included range of outcomes or cautious linguistic expressions (e.g., more likely). Furthermore, advisees

reported greater interest to seek advice from the uncertain adviser in the future. Other researchers also found that overly precise advice was not more influential than its less precise counterparts (Du et al., 2024; Schultze & Loschelder, 2020). Similar patterns are observed with communication of real-world phenomena. For instance, financial forecasts given in ranges rather than as simple point estimates were preferred and perceived as more accurate, informative and credible by investors (Du et al., 2011). This is also the case when communicating scientific findings. For example, expressing greater uncertainty had either no or only a small negative effect when conveying issues such as Genetically Modified Organisms (GMO) food labelling or immigration (Gustafson & Rice, 2019; Van Der Bles et al., 2020). Additionally, acknowledging uncertainty when reporting climate change findings led to greater public acceptance (Joslyn & LeClerc, 2016; Howe et al., 2019). These findings demonstrate that greater confidence is not always preferred. This may be because over precision and confidence reduces the perceived competence of the speaker (Loschelder et al., 2016). In the previously discussed contexts, there are inherent uncertainties, and the outcomes are highly variable. Therefore, highly confident advisers would appear less accurate which leads to lower credibility and advice adoption.

#### **1.6.6. Flexible use of confidence and its link to other factors.**

Past literature suggests that confidence is flexibly used to guide decision-making and behaviour. For instance, confidence does not only represent one's belief regarding how correct an opinion is, but also influences information processing and guide decision-making. One's confidence is not static but dynamically changes based on the accumulation and assessment of evidence (Desender et al., 2021). Additionally, confidence can guide learning of external information. Guggenmos and colleagues (2016)

found that people formulate a current level of confidence and long-term expected confidence. The long-term expected confidence can simulate feedback and enable reinforcement learning. People also use confidence to evaluate difficulty of a task and decide how to tackle problems, such as which task to tackle first (Aguilar-Lleyda et al., 2020).

Additionally, during advice-taking, advisee's confidence is determined by a multitude of factors, such as number of advisers, number of external pieces of information and variability in advisers' answer (Budescu & Rantilla, 2000). Furthermore, advisees are sensitive to subtle differences when evaluating these factors, such as degree of overlap in information accessed by the advisers (Budescu & Rantilla, 2000). People also evaluate advice in relation to their own opinion, where people perceive advice further away from their own opinion to be of lower quality (Meshi et al., 2012) and discount it more than close advice (Molleman et al., 2020; Newell et al., 2009). This effect is amplified when advisees are confident in their own answer (Pescetelli & Yeung, 2021). In addition, advisees use advice not only to increase their decision accuracy but also to test their current accuracy. Schrah and colleagues (2006) found that advisees requested advice after completing 75% of their own information search, indicating advice is used to test their hypotheses rather than to help formulate them. In addition to using advice to evaluate their own opinion, it is also used to evaluate the quality of advice. Pescetelli and Yeung (2021) showed that advisees use their confidence to deduce the adviser's level of calibration between their confidence and accuracy. This evaluation was carried out even in the absence of objective feedback and was nearly as accurate as advisees who received feedback. Additionally, advisees use their confidence to decide whether to request advice and regulate their degree of advice adoption (Pescetelli et al., 2021). However, when

adviser accuracy information is available, advice request better correlated with decision uncertainty, demonstrating advisees use of confidence flexibly switch between adviser evaluation and advice request (Carlebach & Yeung, 2023).

Advisers also express their confidence flexibly during advice-giving. As previously discussed, advisers may overstate their confidence as persuasion (Van Swol, 2009) or helping (Haran et al., 2022) intention increases. However, advisers also express lower confidence when there is higher uncertainty associated with the decision (Haran et al., 2022). Furthermore, Benjamin and Budescu (2014) compared advice given after learning the payoffs of lotteries versus after playing the lottery (i.e., receiving description vs. experiencing). In the latter case, advisers were aware that they did not have the full payoff information, which led them to be more cautious in their recommendation and were more risk averse (e.g., signalling they are unaware of the full payoff). Additionally, advisers who are less confident do not want their advice to be fully adopted and perceive advisees who overweight their advice negatively (Ache et al., 2020). This suggests that reporting lower confidence is not simply a reflection of internal confidence but an attempt to dissuade advisees from adopting their advice. Overall, advisers tend alter their confidence to better reflect the advice-giving context and to manipulate advice adoption.

The above discussed findings demonstrate that confidence is not simply a measure of expectation of being correct but is also related to a multitude of factors. The aim of the current thesis was to examine if advisees are sensitive to these subtle features of the adviser's confidence. Some previous studies have found evidence that advisees make inferences regarding adviser confidence. As previously mentioned, one of the most prominent assumptions of adviser confidence is its relation to the quality of information

and knowledge held by the adviser and thus, their accuracy (i.e., confidence heuristic) (Price & Stone 2004; Thomas & McFadyen, 1995). Similarly, advisees presume that confident advisers would also have higher general expertise (Sniezek & Van Swol, 2001). In support of this, Wang and colleagues (2025) found that advice coupled with high confidence was more influential when it came from advisers with higher task expertise. Researchers also suggest that advisers' confidence also shape perceived trustworthiness of the adviser (Sniezek & Van Swol, 2001). This may be because perception of ability and trust are closely related according to the three-factor model of trust (Mayer et al., 1995). More specifically, high confidence may lead to greater perception of ability which increases trust. Additionally, advisers' high levels of confidence are perceived as greater commitment to their recommendation (Vulllioud et al., 2017). Furthermore, advisees also seem to process adviser confidence in relation to the advice-taking context. As previously mentioned, some recent studies found that uncertain advice receives positive reaction from advisees (e.g., Du et al., 2011; Du et al., 2024; Gaertig & Simmons, 2018; Schultze & Loschelder, 2020). This positive attitude was generally found when there were some levels of uncertainties in the advice-taking context, where high confidence may appear overconfident and less accurate. This suggests that advisees do not process adviser confidence in isolation but in relation to the advice-taking context. However, there is limited research that directly test the relationship between adviser confidence and the advice-taking context. Therefore, empirical work in the current thesis will investigate how advisees use adviser confidence to make inferences about the advice-taking context and features of the adviser.

## 1.7 Thesis Outline

The current chapter reviewed relevant concepts and research in the advice-taking area. First, unique features of advice compared to other forms of social influence was discussed. The research methods in the advice-taking literature were also reviewed, such as task structure (e.g., JAS) and measures of advice adoption (e.g., WOA, WOE). The chapter then reviewed key terms (e.g., egocentric discounting) and factors that are researched in the advice giving and taking literature. The main factors of interest in the current thesis, confidence and expertise, were discussed in greater detail. Past research on adviser confidence and expertise have focused on its influence on single decisions (e.g., advice adoption). However, some researchers suggest that other factors (e.g., accuracy, external cues) influence advisees' processing of adviser confidence. Therefore, the current thesis aimed to examine the information advisees extract from adviser confidence and its relation to the advice-taking context, such as adviser competence (Chapter 2) and the decision-making strategy (Chapter 3 & 4). In line with previous literature, the empirical chapters will examine the role of adviser confidence and expertise on advice-taking but also investigate advisees' flexible use of adviser confidence and expertise.

The aim of Chapter 2 was to answer the following research question: How does the relationship between the advice-taking context and the adviser's linguistic expressions of confidence influence advice-taking? Five experiments investigated how people process different linguistic expressions of confidence in relation to the type of uncertainty in the advice-taking context. As previously mentioned, some past studies showed that advisees sometimes prefer uncertain advice (Du et al., 2011; Du et al., 2024; Gaertig & Simmons, 2018; Schultze & Loschelder, 2020). This preference appears to be related to the level of uncertainty in the advice-taking context. Chapter 2 extends this idea and examined if

advisees are sensitive to the type of uncertainty prevalent in the advice-taking context. In particular, two types of uncertainty proposed by Ülkümen and colleagues (2016) were investigated: epistemic and aleatory uncertainty. The researchers found that people prefer to use confidence terms (e.g., certain, sure, confident) to describe epistemic uncertainty whereas likelihood terms (e.g., probability, chance, likelihood) are favoured to describe aleatory uncertainty (Ülkümen et al., 2016). Chapter 2 examined if this preference for context-appropriate linguistic expressions also translates into advice-taking, where adviser's linguistic expressions influence adviser evaluation and advice-adoption. It was predicted that advisees would be sensitive to the type of uncertainty prevalent in the advice-taking context and prefer advisers who use linguistic expressions that matches the type of contextual uncertainty because it signals the advisers' understanding of the advice-taking context. More specifically, it was hypothesised that when there are high levels of epistemic uncertainty in the advice-taking context, an adviser using confidence term would be more influential and perceived as more competent compared to an adviser using likelihood terms. On the other hand, an adviser using likelihood term would be more influential and perceived as more competent than an adviser using confidence terms when there are high levels of aleatory uncertainty. However, only limited evidence was found across the five experiments.

Chapter 3 reports three experiments investigating how adviser confidence is processed in relation to the available external information. More specifically, the aim was to understand how advisees construe a discrepancy between adviser confidence and external information ambiguity. For example, an adviser could display high levels of confidence despite ambiguous external information or express uncertainty even when the external information seems unambiguous. The effect of this discrepancy on advice adoption was

compared for expert versus novice adviser. It was hypothesised that in the first scenario, participants would increase their confidence when an adviser agreed but decreased under disagreement, where this change in confidence would be greater following the expert's advice. In the second case, it was hypothesised that participants would decrease their confidence despite agreeing advice. This pattern was observed across the three experiments. Additionally, an exploratory hypothesis proposed that participants would change their use of the four basketball statistics following repeated interaction with the advisers. Indeed, participants showed signs of learning and adopting the advisers' decision-making strategy.

Chapter 3 found hints of learning from adviser confidence, but this was not the main purpose of the experiments. Therefore, experiments in Chapter 4 investigated if advisees can learn the validity of external information through adviser confidence. It is possible that the unexpected adviser confidence provides hints about the external information prioritised by the adviser. For example, imagine a situation where two pieces of information supports outcome A, but the remaining two pieces of information suggests outcome B. If an adviser confidently advocates answer A, it is safe to infer that the first two pieces of information are prioritised by the adviser. This was investigated using a basketball game outcome prediction task where participants were given four pieces of basketball statistics. It was hypothesised that advisees would be able to deduce the which basketball statistics are prioritised by the adviser. Additionally, advisees would shift their use of the basketball statistics to mirror the adviser's strategy, which would increase participants accuracy. However, no evidence to support these hypotheses was found across two experiments.

In the final chapter, Chapter 5, the experimental findings are integrated and discussed in relation to the existing advice-taking literature. The findings from the empirical chapters suggest that there might be an effect of advice-taking context on advice-taking under specific conditions. Therefore, future studies could investigate when and how advisees attend to external factors during advice-taking.

The experiments in the empirical chapters followed JAS where participants' answer, and confidence was recorded pre- and post-advice. In each experiment, participants were instructed that an adviser would appear upon submission of their initial answer and provide advice. Minimal instructions were given about the advice to prevent participants from altering how they engage with the task and focusing on advisers rather than the task at hand. This was following standard practice in the advice-taking literature where social information is presented as 'advice' (e.g., Çelen et al., 2010; Hertz et al., 2021; Soll & Larrick, 2009; Yaniv & Kleinberg, 2000). Furthermore, Milyavsky & Gvili (2024) found that labelling information as advice (compared to answers from previous participant) led people to perceive the information as advice even though the helpfulness of the information was identical. Therefore, it was expected that participants would treat the information in the current experiments as advice. However, it is important to note that there are various types of advice (e.g., guidance on decision strategy, socio-emotional support) which may be influenced by various factors (e.g., perceived helpfulness, social relationship) and the current thesis focused on one type of advice: recommendation in favour of a decision. Therefore, the current research findings may only generalise to specific types of advice-taking in the real-world.

# Chapter 2: Adviser Confidence and Uncertainty in Advice-Taking Context

---

## 2.1 Introduction

People receive advice from a wide variety of sources ranging from experts such as doctors, lawyers, and financial advisers, to advice from close others. Taking advice into account during decision-making can improve one's accuracy (Harvey & Fisher, 1997) and result in positive behaviour change (Rose et al., 2013). The benefits of soliciting advice have been clearly demonstrated in numerous studies (Bonaccio & Dalal, 2006), but people do not always effectively incorporate advice into their decision-making (Bailey et al., 2022; Soll & Larrick, 2009). Therefore, researchers have been investigating factors that lead to greater advice-taking. Studies have found factors such as advisee's similarity with the adviser (Chan et al., 2017; Yaniv et al., 2011), adviser's level of expertise (Harvey & Fisher, 1997), trust in the adviser (Sniezek & Van Swol, 2001; Van Swol, & Sniezek, 2005), task difficulty (Schrah et al., 2006), and positive valence (De Hooze et al., 2014) to influence advice adherence. Although various factors influence advice-taking, the current chapter focused on the role of confidence due to its significant impact on advice adoption. As discussed in Chapter 1, the definition of confidence varies across disciplines. However, in the current thesis, confidence is defined as expectation of the extent to which a decision, opinion, or recommendation is correct (Klayman et al., 1999). Adviser confidence is a feature that has consistently been found to have a significant impact on advice-taking. For instance, Price and Stone (2004) showed that advisees prefer overly confident advisers over advisers whose accuracy and confidence are well-calibrated. This preference for confident advice was found across different tasks (e.g., face recognition, pairing identical geometric shapes) and communication medium (e.g., face-to-face, in

writing) (Pulford et al., 2018). In addition, Van Swol and Sniezek (2005) compared five factors that may influence advice adoption and found that only adviser confidence had a significant impact on advice adoption. Price and Stone (2004) argue that the preference for confident advice is because people use adviser confidence as a proxy for how competent the adviser is and deem confident advisers to hold more reliable information (i.e., *confidence heuristic*).

Due to strong evidence demonstrating the power of confident advice, one might assume that people are averse to advice that expresses uncertainty. However, further research suggests that confidence is not what advice-seekers value, but rather what confidence represents – competency. Therefore, when information on adviser accuracy is available, advisees shift their focus from simple confidence to the relationship between confidence and accuracy, and adhere to well-calibrated advisers whose high levels of confidence are warranted (Tenney et al., 2008). Thus, advisers lose credibility when they lack accuracy but display high confidence (Rader et al., 2017; Sah et al., 2013). In addition, research shows that a highly confident adviser is not always preferred. For example, Gaertig and Simmons (2018) found in eight studies that advisees are not averse to and even prefer advice that portrays uncertainty by using probabilistic terms or advisers who give a range of outcomes over exact and precise advice. For example, during a Major League Baseball (MLB) game outcome prediction task, a certain adviser said: “The two teams will accumulate more than 15.5 hits in this game”. On the other hand, uncertain advice made a probabilistic prediction: “There is about a 60% chance that the two teams will accumulate more than 15.5 hits in this game”. Additionally, Du and colleagues (2011) also found that investors prefer financial forecasts given in ranges rather than simple point

estimates. Recent studies have also showed no difference in advice adoption between precise and less precise advice (Du et al., 2024; Schultze & Loschelder, 2020).

This positive attitude towards uncertain advice may be explained by examining the advice-taking context. For example, in Gaertig and Simmons (2018)'s studies, participants predicted future sports game outcomes, weather, and stock prices. The outcomes of these are affected by numerous factors and are highly variable. Therefore, advice that expresses some level of uncertainty appears context appropriate, whereas an overly confident or precise adviser seems overconfident. This suggests that people's preference for confident advisers may depend on the level of uncertainty in the advice-taking context. In other words, people are not opposed to a lack of confidence but want uncertainties to be conveyed in a context-appropriate manner. These results suggest that advisees are sensitive to the level of uncertainty in the advice-taking context. The experiments in the current chapter extend this idea to examine whether advisees are also sensitive to the *type* of uncertainty prevalent in the advice-taking context.

Researchers suggest that there are different types of uncertainties. For example, Kahneman and Tversky (1982) distinguished between internal uncertainty (i.e., stems from lack of one's own knowledge) and external uncertainty (i.e., due to randomness and complexity of causal systems in the environment). A more recent study by Ülkümen and colleagues also (2016) suggest that there are two types of uncertainties: epistemic and aleatory, which were the uncertainty dimensions adopted in the current experiments. *Epistemic uncertainty* refers to uncertainty that arises from lack of knowledge. Thus, the uncertainty can be resolved through information-seeking and by obtaining further knowledge. There is an absolute true or false answer, such as math problems or trivia

questions. Therefore, statements expressing level of *confidence* in one’s knowledge through words, such as *certain*, *sure*, *confident*, are appropriate to describe epistemic uncertainties (Ülkümen et al., 2016). On the other hand, *aleatory uncertainty* characterises uncertainties that arise from randomness in the environment and entails probabilistic outcomes. Therefore, situations with high levels of aleatory uncertainty are unpredictable and should be described in relative frequencies (Ülkümen et al., 2016). For example, it is not possible to be certain about which face will show up on a dice role. The likelihood of obtaining a certain number can only be expressed in probabilistic terms. Therefore, aleatory uncertainties described in *likelihood* terms (i.e., *probability*, *chance*, *likelihood*) are deemed more suitable. The key differences between the two types of uncertainties are illustrated in Table 2.1.

**Table 2.1** *Key differences between epistemic and aleatory uncertainty*

	<b>Epistemic (Knowable) Uncertainty</b>	<b>Aleatory (Random) Uncertainty</b>
Attribution of Uncertainty	Inadequate knowledge/skill	Stochastic behavior
Representation	Single case	Class of possible events
Focus of Prediction	Binary truth value	Tendency to occur
Probability Interpretation	Confidence or causal propensity	Relative frequency
<i>Hypothesized Linguistic Marker</i>	<i>Confidence statements</i>	<i>Likelihood statements</i>

*Note.* Adopted from “Two Dimensions of Subjective Uncertainty: Clues From Natural Language” by G. Ülkümen, C. R. Fox, and B. F. Malle, 2016, *Journal of Experimental Psychology: General*, 145(10), p. 1282. (<http://dx.doi.org/10.1037/xge00002021280>). Copyright 2016 American Psychological Association.

A series of studies conducted by Ülkümen and colleagues (2016) showed that people naturally discriminate between the two types of uncertainties and associate different linguistic expressions with each type of uncertainty. More specifically, participants were

given different scenarios and asked to rate the level of epistemic and aleatory uncertainty in each scenario. Participants consistently rated scenarios to have varying levels of epistemic and aleatory uncertainties. Furthermore, participants preferred to use confidence statements (e.g., “I am X% sure that I will earn at least a 3.0 GPA this semester.”) over likelihood statements (e.g., “I think there is a X% chance that I will earn at least a 3.0 GPA this semester.”) to describe scenarios with high levels of epistemic uncertainty. In contrast, participants’ perceived likelihood statements were more fitting than confidence statements to describe scenarios with high levels of aleatory uncertainty (Ülkümen et al., 2016). Furthermore, people also seem to alter their behaviour in line with their perceived type of contextual uncertainty. For example, people who view the stock market as predictable and high in epistemic uncertainty are more likely to seek financial advice, whereas people who believe the stock market is random (i.e., high in aleatory uncertainty) try to minimise uncertainty by diversifying their investment (e.g., investing in a greater number of stocks) (Walters et al., 2023).

These findings suggest that people distinguish between epistemic and aleatory uncertainty and behave accordingly. Therefore, it is also possible that people expect advisers to also change their advice-giving depending on the advice-taking context. In addition, advisees may also attempt to infer adviser competence from linguistic expressions used in advice. The confidence heuristic argue that advisees deduce adviser competence from adviser confidence. Recent findings suggest that people do not simply equate high confidence as greater competence but are sensitive to subtle cues of competence in adviser confidence such as the level of uncertainty in the advice-taking context. This suggests that advisees process adviser confidence and associated competence in relation to the advice-taking context. Thus, it is also possible that advisees infer advisers who use context-appropriate

linguistic expressions as more competent. This might be because the adviser's linguistic expressions signal their awareness of the type of uncertainty present in the advice-giving context. This would in turn also suggest that the adviser understands how to resolve the uncertainty – either by collecting further information or assessing the randomness. Hence, the adviser's context-appropriate linguistic expressions could increase perceived competence and lead to greater advice adoption. Therefore, the following hypotheses are proposed:

H1: An adviser using confidence terms would be a) perceived as more competent and b) be more influential when there are high levels of epistemic uncertainty than aleatory uncertainty in the advice-taking context

H2: An adviser using likelihood terms would be a) perceived as more competent and b) be more influential when there are high levels of aleatory uncertainty than epistemic uncertainty in the advice-taking context

### **2.1.1 Research Outline**

In order to test the hypotheses, the type of uncertainty (i.e., epistemic vs. aleatory) in the advice-taking context and the advisers' linguistic expressions of confidence were manipulated. Each experiment included two tasks, one designed to have high levels of epistemic uncertainty and one designed to have high levels of aleatory uncertainty. Each task was performed with two advisers, one who used confidence terms (e.g., confident, sure, and certain) to give their advice and another who gave their answer in likelihood terms (e.g., chance, probability, and likelihood). The experiments followed a mixed design where participants only performed one of the two tasks, but all participants

interacted with both advisers. The aim of the experiments was to examine if people prefer advice that conveys an understanding of the type of contextual uncertainty prevalent in the advice-taking context. However, across five experiments, participants showed limited preference for advisers who use context-appropriate linguistic expressions.

## **2.2 Experiment 1**

Experiment 1 examined how difference types of contextual uncertainty (i.e., epistemic vs. aleatory) influence advice-takers' preference for advisers who use confidence versus likelihood statements. Past research showed that people prefer to use confidence terms to express epistemic uncertainties and find likelihood terms are better suited for describing aleatory uncertainties. Therefore, it is possible that people prefer advisers to also use context-appropriate terms. To test this, half of the participants completed a task with high levels of epistemic uncertainty (estimating historic dates) and the remaining participants engaged in a task with high levels of aleatory uncertainty (future temperature prediction). Both advisers were presented in each task, where one adviser gave their answer using confidence terms (e.g., confident, sure, certain) and the other adviser used likelihood terms (e.g., chance, probability, likelihood).

### **2.2.1 Methods**

**Participants.** Participants were recruited through Prolific (<https://www.prolific.com>). In total, 60 participants took part in Experiment 1 (30 = Female, 29 = Male, 1 = Prefer not to say,  $M_{\text{age}} = 32.7$ ,  $SD_{\text{age}} = 10.28$ ). Thirty participated in the historic dates task (i.e., epistemic task) and the remaining 30 participants completed the temperature prediction task (i.e., aleatory task). Prior to engaging in the study, all participants provided informed

consent in accordance with the University of Oxford Central University Research Ethics Committee guidelines. Participants were compensated with £9/hour.

**Task.** The two experimental tasks followed the judge-adviser system (JAS) (Sniezek & Buckley, 1995), a paradigm widely used in the advice-taking research. First, a judge (the participant) gives their initial answer and then views advice from an adviser. After, the judge is given a chance to update their answer. The difference in the judge's initial and final answer reflects the influence of the advice given.

In the *historic dates (epistemic) task*, participants were asked to estimate which year historic events occurred. This task was selected as there is a clear correct answer that is knowable. Participants were given a historic event in bold and an anchor year, and asked if the event occurred before or after the anchor year (e.g., “Did the following event occur before or after the year 1956? **Soviet cosmonaut Yuri Gagarin becomes the first human to travel in space**”). Participants were also asked to provide their accompanying confidence level. Confidence was collected using a horizontal scale, where each end had the two possible answer categories: 100% before to 100% after. The midpoint of the scale corresponded to 50% confident, indicating not selecting either before or after (e.g., in a coin toss 50% chance of heads also means 50% chance of tails). Therefore, the midpoint was unselectable to force an answer. Participants answered by moving a slider along the horizontal scale. Once an initial answer was submitted, one of the two advisers (i.e., either using confidence or likelihood terms) gave their answer and confidence. While viewing the adviser, participants were given a chance to keep the slider at their initial answer or move the slider to change their answer. Here, the slider starting point was participants' initial answer. An example question is shown in Figure 2.1.





dates (epistemic) task, the advisers' answer was formulated by adding a random number to the correct year. Similarly, for the aleatory task, a random number was added to the forecasted temperature, obtained from AccuWeather (<https://www.accuweather.com>), of the relevant city on the date in question. A random number between -10 and +10 was generated from a truncated normal distribution ( $M = 0, SD = 8$ ). Adviser confidence was meaningfully calibrated, where a greater difference in the adviser's answer and anchor year or temperature resulted in greater adviser confidence. For example, if the difference between the adviser's answer and the anchor number was less than or equal to two, the adviser confidence was randomly selected to be an integer between 51 and 60. When the difference was greater than or equal to 10, the advisers expressed confidence between 91 to 100. The advisers' answers were formulated independently from the anchor value. The advisers' answer was not available to the participants and was only used to generate adviser confidence.

**Procedure.** The experiment adopted a mixed design where participants only performed one type of task. However, all participants viewed both the adviser using confidence terms (epistemic adviser) and the adviser using likelihood terms (aleatory adviser) regardless of the task type. The procedure was identical for the historic dates (epistemic) and temperature prediction (aleatory) task.

Participants were first instructed about the task they would perform and were provided with an example question. After providing an initial answer, participants were introduced to the two advisers. One adviser gave advice using words such as *confident*, *sure*, and *certain* (i.e., confidence terms). We termed this adviser as the *epistemic adviser* because previous research has shown that people typically associate these words with epistemic

uncertainty. The other adviser answered using *chance*, *probability*, and *likelihood* terms (i.e., likelihood terms), which people typically associate with aleatory uncertainty, and hence we labelled as the *aleatory adviser*. For instance, the epistemic adviser said, “I am X% confident that...” and the aleatory adviser would state “There is a X% chance that...” (refer to Figure 2.1 and Figure 2.2). In the example question, the two advisers gave identical confidence levels and answers. While viewing the advice, participants gave their final answer to the example question, and then moved on to the training block.

The training block comprised of six trials, where participants viewed each adviser three times. In both tasks, the advisers alternated every trial, and participants viewed each adviser equal number of times. The order the advisers were presented was counterbalanced across participants. After completing the training block, participants completed a test block containing 18 questions in which advisers appeared in an alternating order (same as the training block). Thus, participants viewed each adviser nine times and each term (six terms in total – three confidence, three likelihood) three times during the test block.

Once participants completed the test block, they were asked to rate the competence of each adviser using 7 questions (e.g., *How knowledgeable is this adviser?*, *How credible is this adviser?*) on a 7-point Likert-scale (1 = *Not at all*, 7 = *Extremely*). An overall rating was calculated as the average of these 7 answers. The questionnaire was extracted from Gaertig and Simmons (2018). To investigate if participants could distinguish between the two advisers, an adviser distinction score was collected. More specifically, participants were given six statements involving the six terms (three confidence and three likelihood terms) used by the two advisers and were asked to match the statements to each adviser.

Lastly, participants rated their perceived level of epistemic and aleatory uncertainty of the task using a 4-item Epistemic-Aleatory Rating Scale (EARS), adapted from Walters and colleagues (2023). An answer for each item was collected on a 7-point Likert-scale (1 = *Not at all*, 7 = *Very Much*). Out of the 4-items, two measured the perceived levels of epistemic uncertainty and the remaining two measured the perceived level of aleatory uncertainty. The average of the two items was calculated to represent each type of uncertainty. As previously discussed, epistemic and aleatory uncertainty are two separate dimensions, and a task could have both high levels of epistemic and aleatory uncertainty. All measures are available in the appendices.

**Data Analysis.** The aim of the experiment was to compare how differing levels of epistemic and aleatory uncertainty in the advice-taking context shaped preference for advisers' use of confidence (confident, sure, certain) versus likelihood (chance, probability, likelihood) terms in their advice. The main analyses of interest were therefore of the perceived competence of the two advisers and how influential the advisers were on participants' answers across the two tasks. Thus, a 2 (task type) x 2 (adviser type) ANOVA was conducted. If the interaction effect was significant, pairwise comparisons with Bonferroni correction was carried out. The competency rating of each adviser was compared across the two tasks. In addition to the hypotheses testing, adviser rating was also compared within each task.

Advice influence was measured to examine the degree of advice adoption. Difference in participants' initial and final confidence was used to measure the influence of advice on each trial. This average was calculated separately for cases where the adviser agreed or disagreed with the participants' initial answer. An adviser's overall influence was

calculated as the difference in average confidence change in agreeing and disagreeing trials:

$$\text{Adviser influence} = M\Delta_{\text{confidence for agreement}} - M\Delta_{\text{confidence for disagreement}}$$

A 2 (task type) x 2 (adviser type) ANOVA with adviser influence as the dependent variable was conducted to test if the advisers' relative influence differed by task type. Identical to the ANOVA on adviser competence rating, pairwise comparisons with Bonferroni correction were conducted if the interaction effect was significant. Each adviser's influence was compared across the two tasks to test the hypotheses. Additional pairwise comparison with Bonferroni correction were carried out to compare the influence of the two advisers within each task.

Additional analyses were carried out as checks on the effectiveness of the experimental manipulations. First, the two tasks were designed to have varying levels of epistemic and aleatory uncertainty. Participants were expected to perceive the historic dates task to have high levels of epistemic uncertainty but low levels of aleatory uncertainty. The opposite pattern was expected for the temperature prediction task. To test this, a 2 (task type) x 2 (epistemicness vs. aleatoriness) ANOVA was conducted with the EARS score as the dependent variable. Another analysis assessed whether participants distinguished between the two advisers (adviser distinction score), in terms of identifying which adviser used confidence or likelihood terms. The maximum score for the adviser distinction questions was six (three confidence and three likelihood terms), and thus participants were expected to perform above chance level (i.e., get score above three). Therefore, participants' adviser distinction score was tested to establish if it was significantly different from three using a t-test. Additionally, it was necessary to ensure that the task

difficulty was similar. Therefore, participants' initial accuracy in the historic dates (epistemic) task and the temperature prediction (aleatory) task was compared using a t-test.

### 2.2.2 Results

**Table 2.2** *Descriptive statistics of measures in Experiment 1*

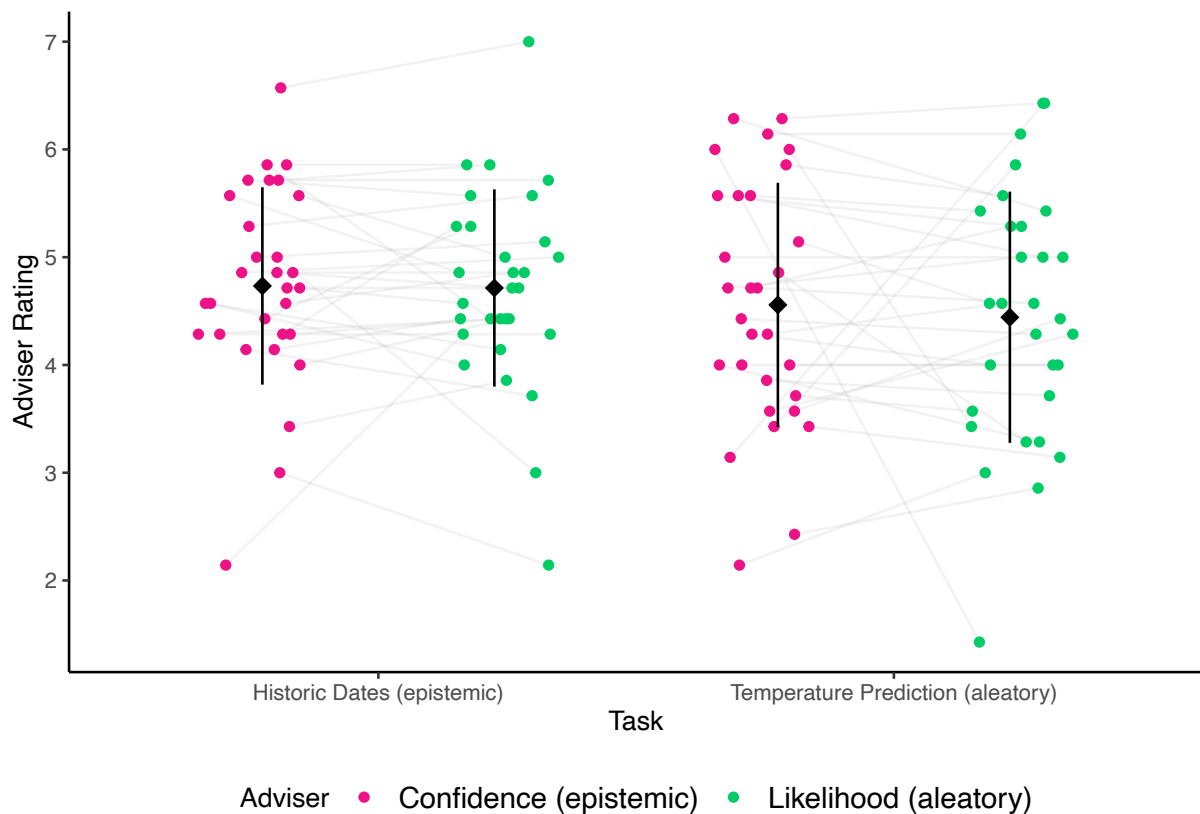
	Historic Dates (epistemic) task		Temperature Prediction (aleatory) task	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Epistemic Adviser Rating	4.7	0.2	4.6	0.2
Aleatory Adviser Rating	4.7	0.2	4.4	0.2
Epistemic Adviser Influence	28.2	3.4	12.7	2.9
Aleatory Adviser Influence	27.7	3.8	12.1	3.4
Task Epistemicness	11.3	0.4	10.3	0.4
Task Aleatoriness	5.8	0.5	8.2	0.6
Adviser Distinction Score	2.7 (2.0)		3.7 (2.0)	

*Note.* Values in parentheses are standard deviations.

**Adviser Rating.** Adviser rating was analysed to investigate if there was a difference in perceived adviser competence by task type. A 2 (task type) x 2 (adviser type) ANOVA was carried out to inspect if there is a difference in adviser competence rating by task type. There was no main effect of task type nor adviser type, where both  $F < 1$ , ( $\eta^2 = .012$ ,  $\eta^2 = .001$ , respectively). The adviser rating was not expected to differ based on task type or overall be particularly higher for one adviser. However, an interaction effect was expected as it was hypothesised that the adviser using confidence terms (epistemic adviser) would be rated as more competent in the historic dates (epistemic) task than in the temperature prediction (aleatory) task. On the other hand, the likelihood term-using (aleatory) adviser would be perceived as more competent in the temperature prediction (aleatory) task than in the historic dates (epistemic) task. The interaction effect between task type and adviser type was not significant,  $F < 1$ ,  $\eta^2 < .001$ . As seen in Figure 2.3,

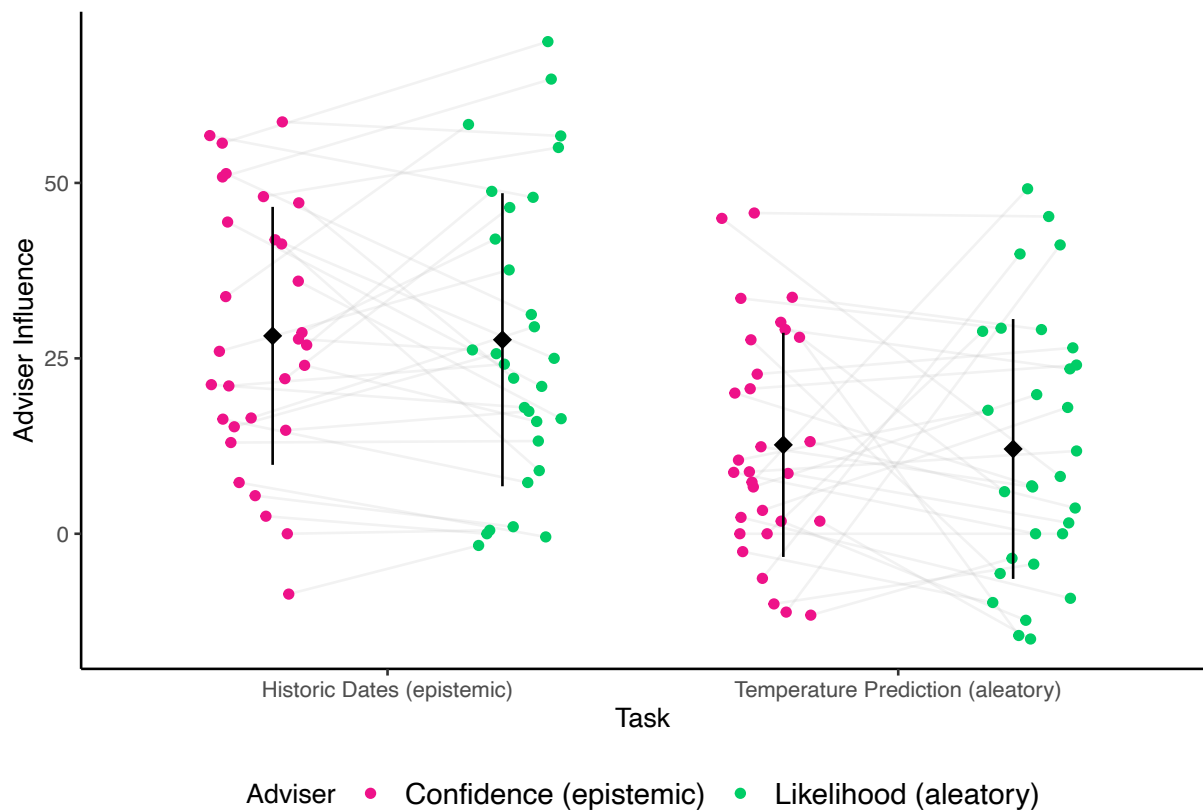
adviser competency rating was similar for both tasks and advisers. The lack of significant effects was unexpected. Therefore, a Bayesian ANOVA was conducted to compare the fit of the data between the null and alternative hypotheses using the *BayesFactor* package in R (Morey & Rouder, 2018) with the default Cauchy prior,  $r = 0.707$ . However, the results were consistent regardless of the prior used. The analysis found anecdotal evidence against the main effect of task type ( $BF_{10} = 0.43$ ) and substantial evidence against adviser type ( $BF_{10} = 0.22$ ). Of critical interest was the interaction effect. Therefore, the model with the two main effects and the full model (two models and interaction term) was compared, which found substantial evidence for the absence of interaction effect ( $BF_{10} = 0.28$ ). Therefore, no evidence was found to suggest that the interaction term meaningfully explain the variation in the data.

**Figure 2.3** *Adviser rating by task type in Experiment 1*



**Adviser Influence.** The second main analysis was regarding adviser influence. The influence of advice was measured as participants' change in confidence from initial to final answer as a function of whether the adviser agreed or disagreed with participants' initial decision. A 2 (task type) x 2 (adviser type) ANOVA with adviser influence as the dependent measure was conducted. There was a significant main effect of task type,  $F(1, 58) = 14.23, p < .001, \eta^2 = .15$ . Both advisers were more influential in the historic dates (epistemic) task than in the temperature prediction (aleatory) task (Figure 2.4). The inherent randomness in the aleatory task might have reduced the perceived usefulness of advice and reduced the overall adviser influence. There was no main effect of adviser type,  $F < 1, \eta^2 < .001$ . The interaction effect was examined to investigate if each adviser's influence differed by task type (H1 & H2). A significant interaction effect and pairwise comparison would demonstrate that the adviser using confidence term is more influential in the historic dates (epistemic) task than in the temperature prediction (aleatory) task. On the other hand, the likelihood term using adviser would be more influential in the temperature prediction (aleatory) task than in the historic dates (epistemic) task. Contrary to the hypotheses, there was no significant interaction effect between task type and adviser type,  $F < 1, \eta^2 < .001$ . Same results were obtained with Bayesian ANOVA, which found decisive evidence for a main effect of task type ( $BF_{10} = 71.54$ ). However, there was substantial evidence against the main effect of adviser type ( $BF_{10} = 0.20$ ). The Bayes Factor for the interaction term was calculated by comparing the main effects only model and full model (main effects and interaction), which found decisive evidence for the absence interaction effect ( $BF_{10} = 0.26$ ).

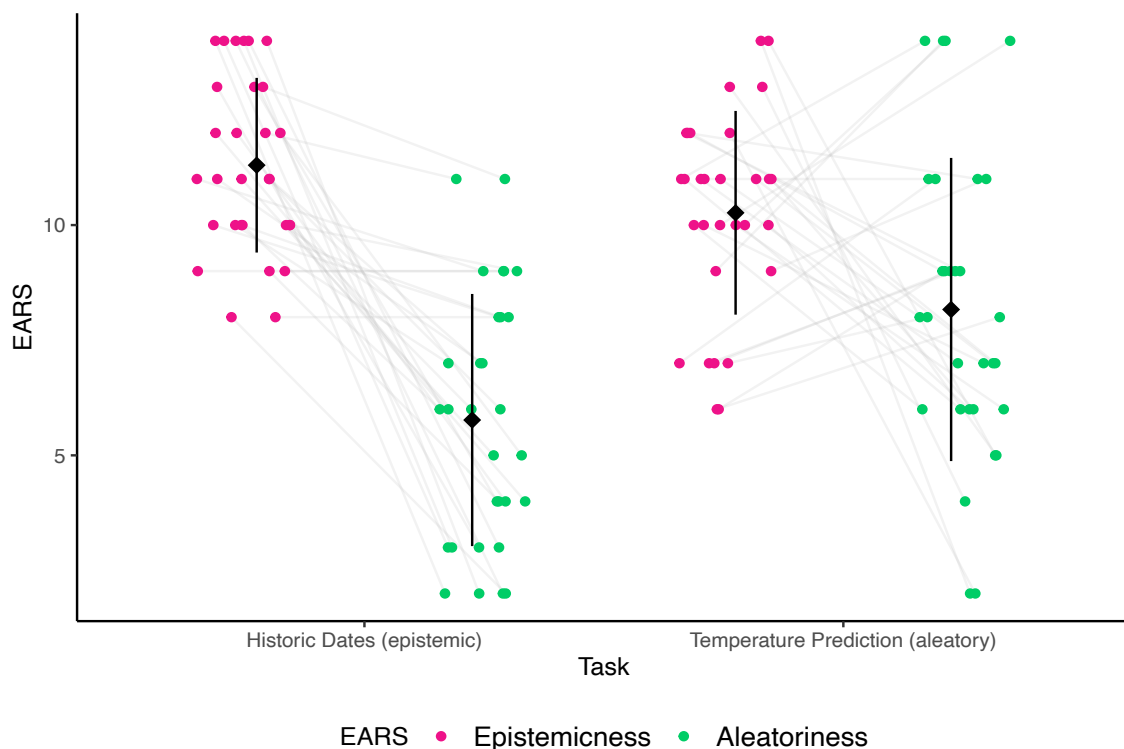
**Figure 2.4** *Adviser influence by task type in Experiment 1*



***Epistemic-Aleatory Rating Scale (EARS)***. Participants' perceived level of epistemic and aleatory uncertainty of each task was recorded. A 2 (task type) x 2 (epistemicness vs. aleatoriness) ANOVA was conducted with overall rating as the dependent variable. The main effect of task type was not significant,  $F(1, 58) = 3.45, p = .068, \eta^2 = .02$ . This was expected as the historic dates (epistemic) task would score highly for epistemicness but low for aleatoriness. In contrast, the opposite would be true for temperature prediction (aleatory) task and thus, the overall rating would be similar. On the other hand, there was a significant main effect of rating type,  $F(1, 58) = 46.76, p < .001, \eta^2 = .36$ . The overall epistemicness rating was greater than the overall aleatoriness rating. This was driven by the temperature prediction (aleatory) task where the epistemicness and aleatoriness rating were both high. The interaction effect was the main focus of the current analysis as it was expected that perceived epistemic uncertainty would be higher in the historic dates (epistemic) task whereas aleatoriness rating would be greater in the temperature

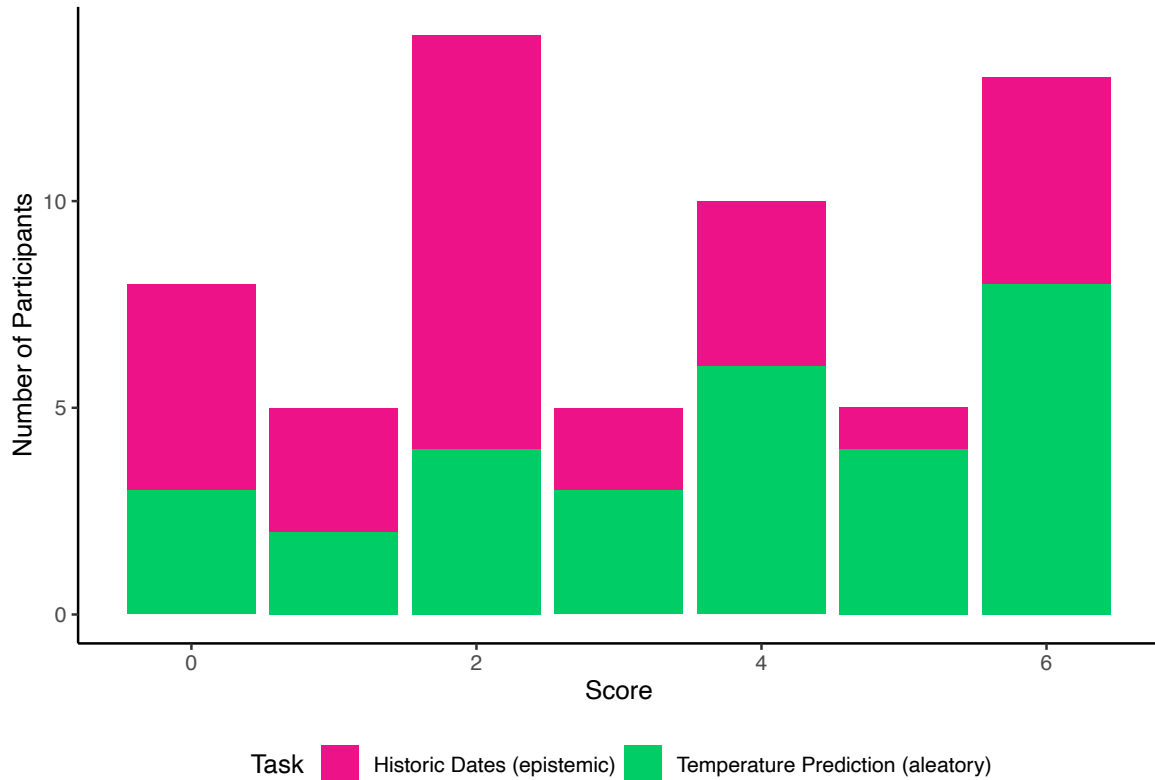
prediction (aleatory) task. There was a significant interaction effect between task and rating types (epistemicness vs. aleatoriness),  $F(1, 58) = 9.46, p = .003, \eta^2 = .10$ . Therefore, a pairwise comparison with Bonferroni correction was carried out to further examine the interaction effect. The perceived level of epistemic uncertainty did not differ significantly by task type,  $p = .057, d = -0.50$ . However, the epistemicness rating was numerically higher in the historic dates (epistemic) task compared to the temperature prediction (aleatory) task (Figure 2.5). As predicted, the aleatoriness rating was significantly higher in the temperature prediction (aleatory) task than in the historic dates (epistemic) task,  $p = .003, d = 0.80$ . For participants performing the historical dates (epistemic) task, their task epistemicness rating was significantly higher than the task aleatoriness rating,  $p < .001, d = -2.35$ . Similarly, participants' epistemicness rating was significantly higher than the aleatoriness rating within the temperature prediction (aleatory) task,  $p = .005, d = -0.75$ . This was unexpected as the temperature prediction (aleatory) task was designed to have high levels of aleatory uncertainty rather than epistemic uncertainty.

**Figure 2.5** Epistemic aleatory rating scale by task type in Experiment 1



**Adviser distinction.** The adviser rating and influence results were not as hypothesised despite the EARS results, which suggested that the task uncertainty manipulation was fairly successful. Therefore, whether participants could distinguish between the two advisers were examined. Analysis of the adviser distinction score was carried out to determine if participants were able to identify which terms (confidence vs. likelihood) were used by each adviser. Score for the adviser distinction was calculated based on the number of statements participants correctly paired with each adviser. The maximum score was six (three confidence and three likelihood terms). Participants' adviser distinction score was not significantly better than chance level ( $M = 3.18$ ,  $SD = 2.06$ ),  $t < 1$ ,  $d = 0.09$ . Furthermore, most participants' adviser distinction score was below three (Figure 2.6). This demonstrates that participants had difficulties identifying the terms used by each adviser.

**Figure 2.6** Adviser distinction score by task type in Experiment 1



**Accuracy.** Participants' initial accuracy (prior to receiving advice) was calculated using the correct year for the historic dates (epistemic) task and estimated future temperature

retrieved from AccuWeather (<https://www.accuweather.com>) for the temperature prediction (aleatory) task. Participants' initial accuracy for the two tasks did not differ significantly,  $t < 1$ ,  $d = 0.18$ . This affirmed that the disparity in task difficulty was not substantial, as intended. Additionally, a 2X2 ANOVA was carried out with time point (initial vs final) and task type (historic dates vs temperature prediction) as the independent variables and accuracy as the dependent variable. The main effect of time point was significant,  $F(1, 58) = 57.18$ ,  $p < .001$ ,  $\eta^2 = .225$ . This suggests that participants accuracy significantly improved following advice (Table 2.3). However, the main effect of task type was not significant  $F < 1$ ,  $p = .512$ ,  $\eta^2 = .005$ . The interaction effect was also not significant  $F < 1$ ,  $p = .760$ ,  $\eta^2 < .001$

**Table 2.3** Participant's Accuracy by task in Experiment 1

	Historic Dates (epistemic) task		Temperature Prediction (aleatory) task	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Initial Accuracy (%)	63	0.15	61	0.12
Final Accuracy (%)	76	0.14	75	0.10

### 2.2.3 Discussion

In the current experiment, it was hypothesised that advisees would prefer advisers who use context appropriate linguistic expressions. Additionally, this preference for advisers to use different linguistic expressions would stem from the assumptions that the adviser has a better understanding of the contextual uncertainty and thus more competent. Contrary to predictions, the results showed that ratings of the two advisers' (confidence vs. likelihood terms) competence and influence did not differ by task type. Additionally, within the historic dates (epistemic) task, the competence rating of the adviser using confidence terms and adviser using likelihood terms did not significantly differ, which was also the case within the temperature prediction (aleatory) task. The two advisers' influence also did not differ across the two tasks nor within each task.

These findings were unanticipated as according to past research, people coherently distinguish between the two types of uncertainties and tend to use linguistic expressions that are consistent with their perceived type of uncertainty (Ülkümen et al., 2016). Thus, it was expected that this would transfer to advice-taking and people would prefer advice given in linguistic expressions that matches the type of contextual uncertainty. The advice-taking context had an effect on overall adviser influence, where both advisers were significantly more influential in the historic dates (epistemic) task rather than in the temperature prediction (aleatory) task. This difference could be due to the nature of the two tasks. Participants might have believed that predicting future temperature has an element of randomness and thus, advice is not useful.

The overall results were not as anticipated therefore the robustness of the methodology was tested by examining the Epistemic-Aleatory Rating Scale (EARS) score and adviser distinction score. As expected, the EARS score showed that participants perceived the historic dates (epistemic) task to be higher in epistemic uncertainty rather than aleatory uncertainty. This suggests that people perceive uncertainty in estimating historic dates to arise primarily from lack of knowledge rather than contextual randomness. The aleatory uncertainty rating for the temperature prediction (aleatory) task was significantly higher than the historic dates (epistemic) task. Therefore, participants viewed predicting future temperature to have a greater element of randomness than guessing historic dates. Surprisingly, the epistemicness rating in the temperature prediction (aleatory) task was significantly higher than the aleatoriness rating. This result was unexpected as past studies used a variation of the temperature prediction task (e.g., estimating weather) and found that participants perceived the task to have high levels of aleatory uncertainty

(Tannenbaum et al., 2017). Despite the similar epistemic and aleatory uncertainty rating within the temperature prediction (aleatory) task, the task uncertainty manipulation was mostly successful and unlikely to explain the lack of findings. Next, the adviser distinction score was examined which was not significantly better than chance level, indicating that participants had difficulties associating each term to the two advisers. It is possible that the results in the current experiment were driven by this difficulty distinguishing between the two advisers. In the experiment, outline figures were used to represent the two advisers and the advisers alternated after each trial. This might have made it challenging to differentiate and form distinct representations of the two advisers.

The methodological shortcomings of Experiment 1 hindered the testing of the hypotheses. Therefore, we ran a follow-up experiment with changes that were designed to facilitate the distinction between the two advisers. More specifically, photos were used to represent the advisers instead of coloured outline figures. Additionally, the same adviser was presented for a few successive trials and then alternated, rather than alternating after every trial. This should allow participants to become accustomed to each adviser. However, the tasks for Experiment 2 remained identical to Experiment 1. The historic dates (epistemic) task was perceived to have higher levels of epistemic uncertainty than aleatory uncertainty. Although the within task uncertainty rating for the temperature prediction (aleatory) task was not as expected, the aleatory uncertainty rating was still higher in the temperature prediction (aleatory) task than in the historic dates (epistemic) task. Therefore, hypotheses testing was still expected to be possible despite some limitations in task manipulation.

## 2.3 Experiment 2

Experiment 2 was designed to improve the methodological flaws of Experiment 1 to better test if the different types of contextual uncertainties affect perception of advisers' linguistic expressions. The adviser distinction score in Experiment 1 showed that participants were unable to consistently distinguish between the two advisers. Therefore, in Experiment 2, advisers were represented using faces of real people, and each adviser appeared for a few trials in a row, to facilitate differentiation of advisers. The same tasks were used in Experiment 2 to reproduce the successful task manipulation in Experiment 1. Once again, participants only performed one task, but all participants viewed two advisers – one using confidence terms and the other using likelihood terms.

### 2.3.1 Methods

**Participants.** In Experiment 2, 60 participants (29 = Female, 28 = Male, 1 = Non-binary, 1 = Prefer not to say,  $M_{\text{age}} = 37.3$   $SD_{\text{age}} = 13.02$ ) were once again recruited through Prolific (<https://www.prolific.com>). There were 30 participants performing the historic dates (epistemic) task and 30 participants in the temperature prediction (aleatory) task. Informed consent was obtained from all participants before engaging in the study and participants were compensated with £9/hour.

**Task.** The tasks in Experiment 2 were identical to those used in Experiment 1. For the task designed to have high levels of epistemic uncertainty, participants estimated if a historic event occurred before or after a specific year (epistemic task). In the second task, participants predicted if the temperature of a given city on a specific future date would be higher or lower than a given temperature (aleatory task).

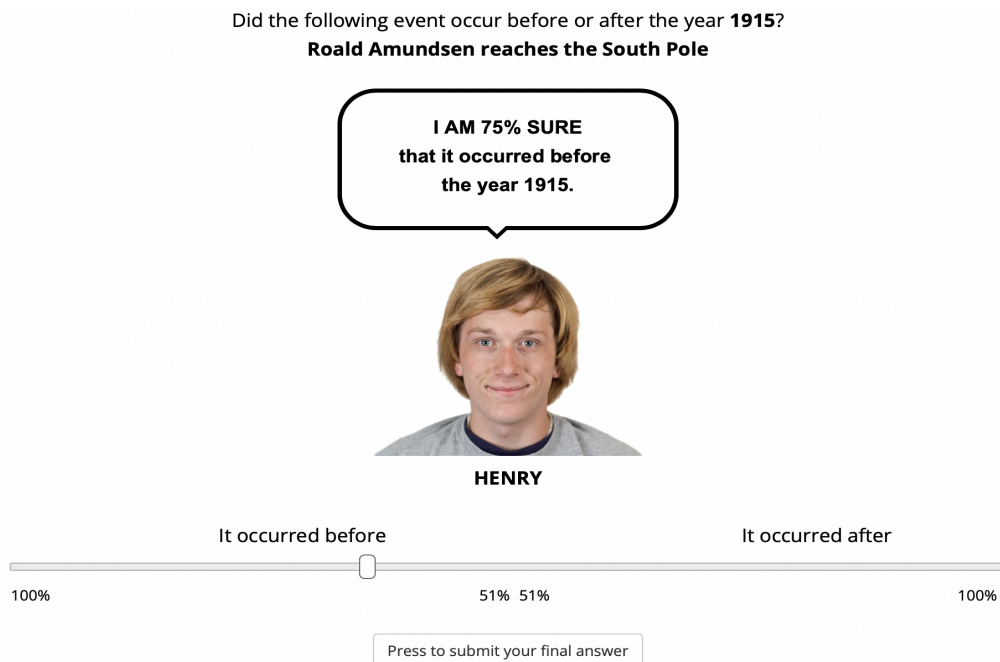
**Advisers.** The adviser distinction score from Experiment 1 showed that participants had difficulties distinguishing between the two advisers. Thus, pictures of real people were selected from the Chicago Face Database (CFD) (<https://www.chicagofaces.org>) on the assumption that these would be more distinctive to participants. Two pictures of white men who were smiling with their mouth closed were chosen to represent the two advisers and were also given names (James and Henry). Additionally, the two faces were chosen to be matched for norm-rated trustworthiness and attractiveness based on the CFD data, as these factors have been found to influence advice adherence (Strong, 1968). Adviser features (e.g., gender, race, trustworthiness) were controlled as these factors could be a source of variance during advice-taking. Since the aim of the experiment was to examine the effect of advisers' expressions of confidence (i.e., confidence vs. likelihood terms), these additional adviser features were controlled to better detect the effects.

In addition to changing the advisers' pictures, the shape of the speech bubbles was different for each adviser (square vs. smooth edges). Furthermore, advice appeared on the screen line-by-line to encourage participants to read all of the advice and the linguistic expressions (i.e., confidence or likelihood terms) were capitalised. The adviser pictures, names and presentation order were randomised. The two advisers are shown in Figures 2.7 and 2.8. Adviser confidence was generated using the same method as Experiment 1. Advice was devised by adding a random number to the correct answer. The difference between the advisers' answer and the anchor year or temperature determined the advisers' confidence, where greater difference led to higher confidence.

**Procedure.** The overall procedure of Experiment 2 was nearly identical to Experiment 1. Participants for both tasks were first given instructions with an example question and

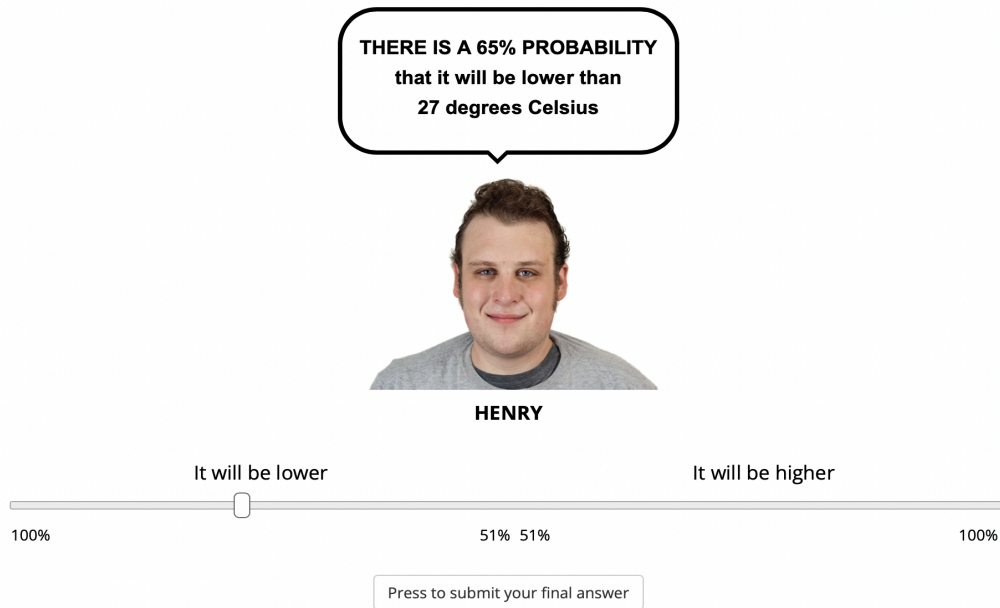
submitted their initial answer and confidence level using the same horizontal scale as Experiment 1. After, they were given advice from the two advisers. The two advisers gave identical answers and confidence level but only differed in the linguistic expressions (i.e., confidence vs. likelihood terms) used in their advice. Participants were given a chance to update their answer while viewing the advisers (i.e., the experimental procedure followed a Judge Adviser System). As in Experiment 1, there were six trials in the training block and 18 trials in the test block. However, in contrast to Experiment 1 in which the advisers alternated every trial, here one adviser appeared for three consecutive trials and then the other adviser was shown for the next three trials. The order each adviser was presented was counterbalanced. After the test block, participants rated each advisers' competence, completed the adviser distinction questions and the Epistemic-Aleatory Rating Scale (EARS) using the same materials as Experiment 1 (refer to appendix A). Examples of the two task questions and advisers are shown in Figures 2.7 and 2.8.

**Figure 2.7** Example question of historic dates (epistemic) task in Experiment 2



**Figure 2.8** Example question of temperature prediction (aleatory) task in Experiment 2

Will the high temperature in New York, U.S.A. on June 1, 2023 be lower or higher than **27** degrees Celsius?



**Data Analysis.**

Experiment 2 remained largely similar to Experiment 1 and thus data were analysed following the same method. In order to ensure that the task manipulation was successful, participants' perceived level of epistemic and aleatory uncertainty in each task was analysed. A 2 (task type) x 2 (epistemicness vs. aleatoriness) ANOVA with the Epistemic-Aleatory Rating Scale (EARS) score as the dependent variable was conducted. The adviser distinction score (max. score six) was tested against three through a t-test. This was to examine if participants can distinguish between the two advisers above chance level. Additionally, to ensure that the two tasks had a similar level of difficulty, participants initial accuracy was compared across the two tasks using a t-test.

The main objective of the experiment was to examine if there was a difference in adviser rating and influence by task type (H1 & H2). Therefore, two 2 (task type) x 2 (adviser type) ANOVAs were conducted, one with adviser rating as the dependent variable and

another with adviser influence as the dependent variable. Adviser influence was calculated following the same method as Experiment 1.

### 2.3.2 Results

**Table 2.4** *Descriptive statistics of measures in Experiment 2*

	Historic Dates (epistemic) task		Temperature Prediction (aleatory) task	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Task Epistemicness	10.9	0.4	8.7	0.4
Task Aleatoriness	5.7	0.5	9.7	0.4
Epistemic Adviser Rating	4.2	0.3	4.1	0.2
Aleatory Adviser Rating	3.6	0.3	4.3	0.2
Epistemic Adviser Influence	27.6	4.2	16.9	3.2
Aleatory Adviser Influence	20.0	3.9	19.4	3.4
Adviser Distinction Score	3.5 (1.8)		3.9 (1.4)	

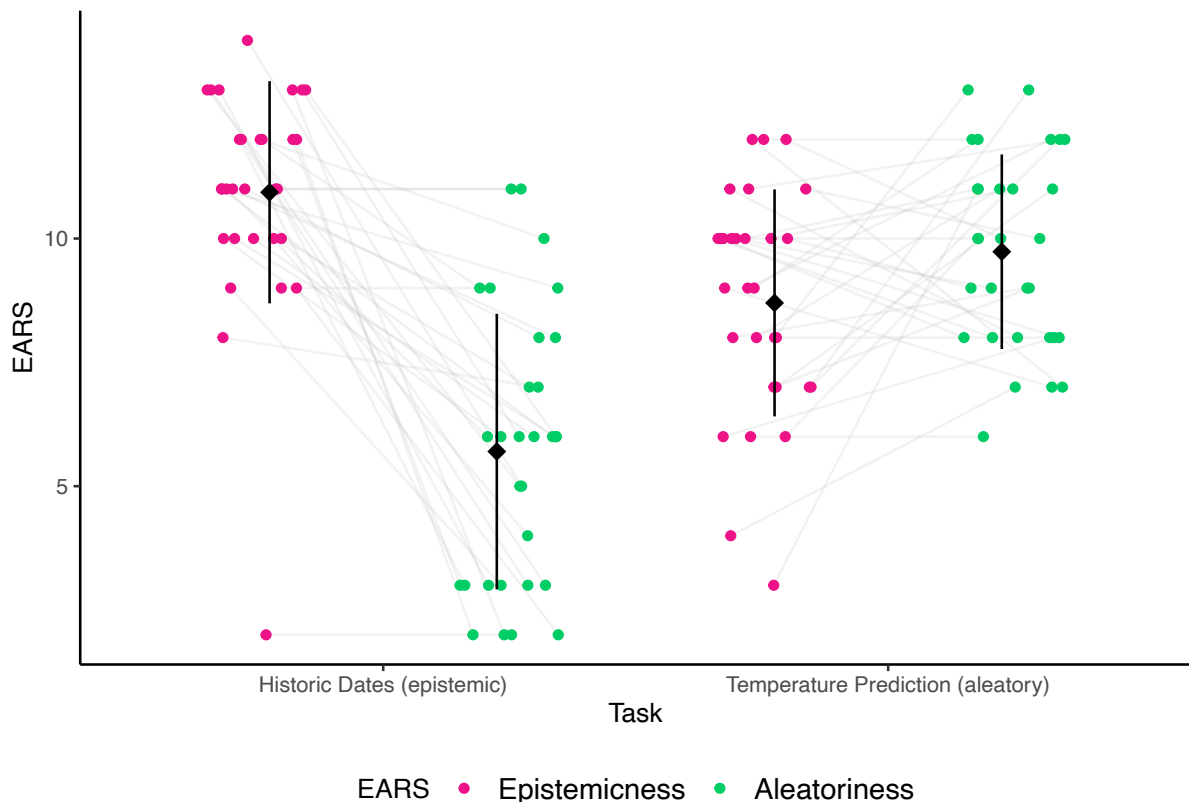
*Note.* Values in parentheses are standard deviations.

***Epistemic-Aleatory Rating Scale (EARS).*** Prior to testing the hypotheses, additional analyses were conducted to ensure that the experimental manipulations were successful. Firstly, analysis of participants' perceived level of epistemic and aleatory uncertainty for the historic dates (epistemic) task and temperature prediction (aleatory) task was conducted. In light of previous literature and Experiment 1, it was expected that the epistemicness rating would be higher in the historic dates (epistemic) task whereas the aleatoriness rating would be greater in the temperature prediction (aleatory) task.

A 2 (task type) x 2 (epistemicness vs. aleatoriness) ANOVA revealed significant main effects of task type,  $F(1, 58) = 4.24, p = .044, \eta^2 = .04$ , and rating type,  $F(1, 58) = 25.40, p < .001, \eta^2 = .17$ . Similar to Experiment 1, the temperature prediction (aleatory) task was rated highly in both aleatory and epistemic uncertainty, thus giving rise to the two significant main effects observed. Of critical interest was the interaction between task type and rating type, which was also significant,  $F(1, 58) = 56.53, p < .001, \eta^2 = .32$ .

Pairwise comparison with Bonferroni correction revealed that perceived level of epistemic uncertainty was greater in the historic dates (epistemic) task compared to the temperature prediction (aleatory) task,  $p < .001$ ,  $d = -1.00$ . On the other hand, perceived level of aleatory uncertainty was higher in the temperature prediction (aleatory) task,  $p < .001$ ,  $d = 1.68$ . Additionally, the epistemicness rating was significantly higher than the aleatoriness rating within the historic dates (epistemic) task,  $p < .001$ ,  $d = -2.07$ . However, as visible in Figure 2.9, within temperature prediction (aleatory) task, the aleatoriness rating was numerically greater than the epistemicness rating, but this difference was not significant,  $p = .066$ ,  $d = 0.50$ .

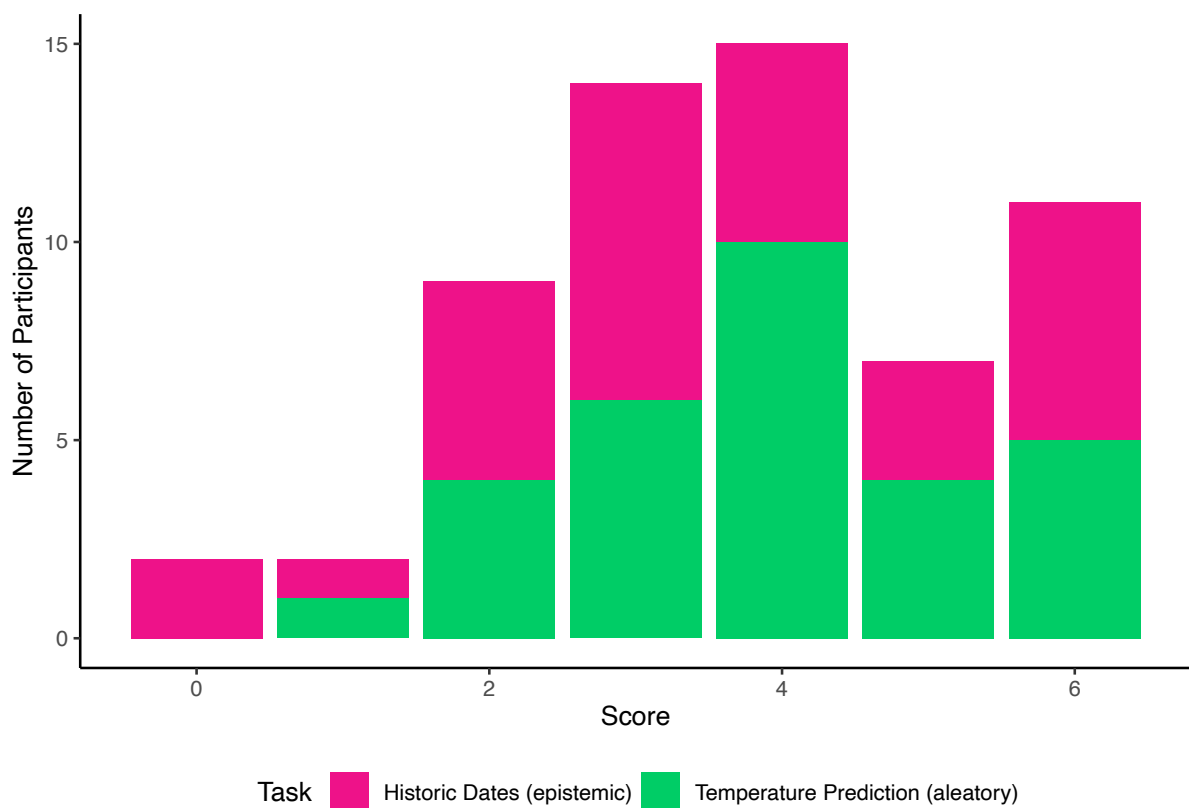
**Figure 2.9** Epistemic-Aleatory Rating Scale (EARS) by task in Experiment 2



**Adviser distinction.** Experiment 1 showed that participants had difficulties distinguishing between the two advisers. This is likely to have contributed to the lack of findings in Experiment 1, and thus, methodological changes were made to facilitate recognition of

each adviser (e.g., pictures to represent advisers). In Experiment 2, participants' adviser distinction score was significantly above chance level ( $M = 3.72$ ,  $SD = 1.57$ ),  $t(59) = 3.53$ ,  $p < .001$ ,  $d = 0.46$ . This finding suggests that participants were able to distinguish between the two advisers. Figure 2.10 shows that the majority of participants scored three or four but also about 20% of participants accurately matched all statements to the correct adviser (i.e., score six).

**Figure 2.10** Adviser distinction score by task type in Experiment 2



**Accuracy.** In order to ensure that task difficulty was at a similar level for both tasks, participants' initial accuracy (accuracy before advice) was calculated following the same method as Experiment 1. Participants' accuracy in the historic dates (epistemic) task did not significantly differ compared to the participants' accuracy in the temperature prediction (aleatory) task,  $t(58) = -1.62$ ,  $p = .110$ ,  $d = -0.42$ . However, participants' initial accuracy was numerically higher in the temperature prediction (aleatory) task. A 2 (initial

vs. final) x 2 (historic dates vs. temperature prediction) ANOVA found a significant main effect of time point,  $F(1, 58) = 40.6, p < .001, \eta^2 = .12$ . However, the main effect of task type  $F(1, 58) = 1.68, p = .20, \eta^2 = .023$ , and the interaction term,  $F(1, 58) = 1.37, p = .247, \eta^2 = .005$ , were not significant.

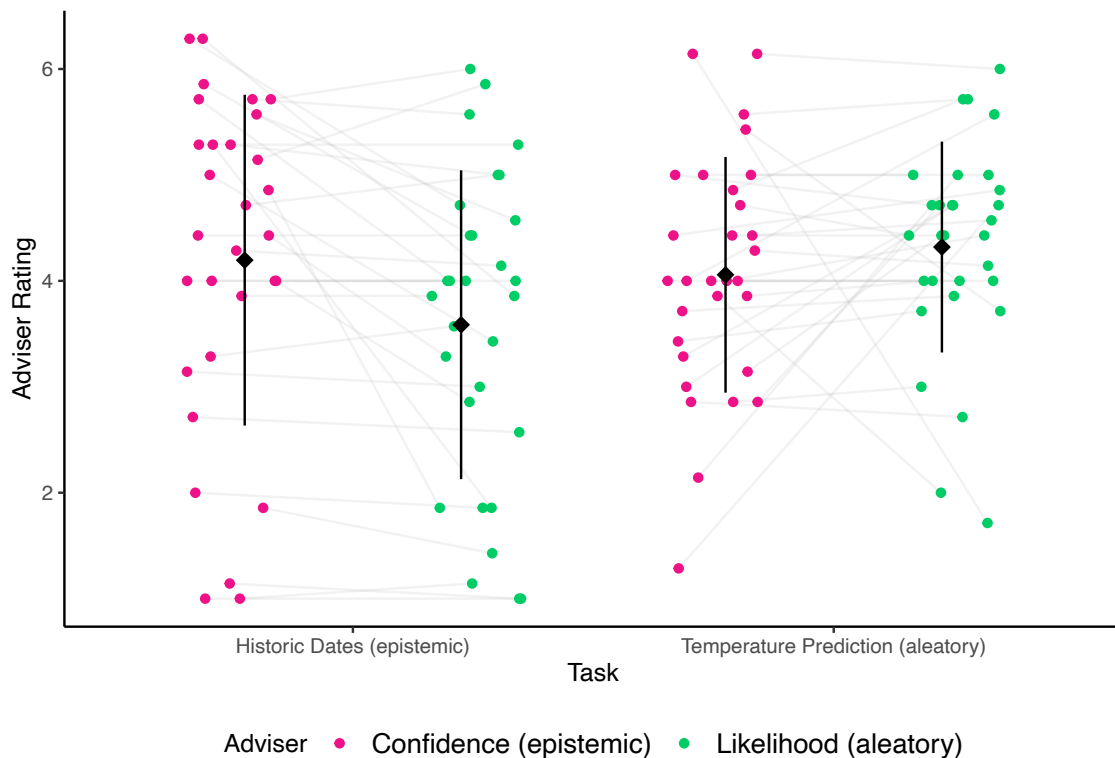
**Table 2.5** Participant's Accuracy by Task in Experiment 2

	Historic Dates (epistemic) task		Temperature Prediction (aleatory) task	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Initial Accuracy (%)	60	0.14	65	0.11
Final Accuracy (%)	71	0.14	73	0.10

**Adviser Rating.** One of the main predictions was that there would be a difference in perceived competence between the two advisers (using confidence vs. likelihood terms) based on task uncertainty (H1 & H2). This was tested with a 2 (task type) x 2 (adviser type) ANOVA. There was no significant main effect of task type,  $F(1, 58) = 1.01, p = .319, \eta^2 = .01$ . The main effect of adviser type was also not significant,  $F(1, 58) = 1.19, p = .279, \eta^2 = .005$ . The main effects were not expected to be significant as the overall adviser rating should be similar between the two tasks. However, critically, there was a significant interaction between task type and adviser type,  $F(1, 58) = 7.50, p = .008, \eta^2 = .03$ . The interaction effects were further examined through pairwise comparisons with Bonferroni correction. The confidence term-using (epistemic) adviser rating did not differ significantly by task,  $p = .695, d = -0.10$ . On the other hand, the likelihood term-using (aleatory) adviser was rated as significantly more competent in the temperature prediction (aleatory) task than in the historic dates (epistemic) task,  $p = .027, d = 0.60$ . Within the historic dates (epistemic) task, the rating of the confidence term-using (epistemic) adviser was numerically greater, but the difference was not significant,  $p = .123, d = -0.40$ . Similarly, within the temperature prediction (aleatory) task, the likelihood term-using (aleatory) adviser rating was numerically higher, but the difference was not significant,  $p$

$= .341, d = 0.25$ . A Bayesian ANOVA was also carried out to examine the support for the hypotheses. There was anecdotal evidence against the main effect of task type ( $BF_{10} = 0.40$ ). The evidence against the main effect of adviser type ( $BF_{10} = 0.25$ ) was substantial. The Bayes Factor for the interaction term was derived by comparing the model that only included the two main effects and the model that also included the interaction term, which found substantial evidence for the interaction term ( $BF_{10} = 5.76$ ).

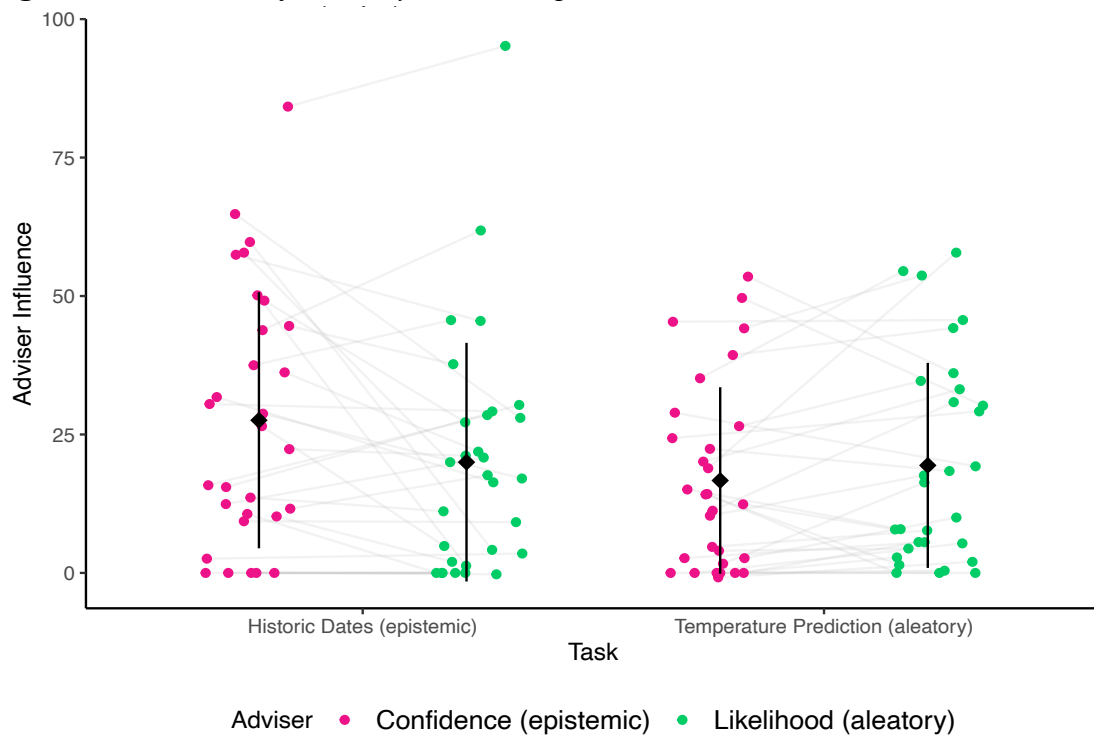
**Figure 2.11** *Adviser Competency Rating by Task in Experiment 2*



**Adviser Influence.** It was also hypothesised that the degree of advice adoption for each adviser would vary by task type (H1 & H2). In order to test the hypotheses, a 2 (task type) x 2 (adviser type) ANOVA was carried out with adviser influence as the dependent variable. The main effect of task type was not significant,  $F(1, 58) = 1.41, p = .239, \eta^2 = .02$ . This was also the case for adviser type,  $F(1, 58) = 1.51, p = .224, \eta^2 = .004$ . Adviser influence was not expected to vary by task or adviser. The main interest was the interaction effect between task type and adviser, which was significant,  $F(1, 58) = 6.83,$

$p = .011$ ,  $\eta^2 = .02$ . The adviser using confidence terms (epistemic adviser) was significantly more influential in the historic dates (epistemic) task compared to the temperature prediction (aleatory) task,  $p = .042$ ,  $d = -0.54$ . The influence of the adviser using likelihood term (aleatory adviser) did not differ significantly by task type,  $p = .912$ ,  $d = -0.03$ . Within the historic dates (epistemic) task, the confidence term-using (epistemic) adviser's influence was numerically greater than the adviser using likelihood terms (aleatory adviser), but the difference was not significant,  $p = .195$ ,  $d = -0.34$ . Adviser influence did not differ significantly within the temperature prediction (aleatory) task,  $p = .553$ ,  $d = 0.20$ . Yet the numerical trend was as predicted with likelihood term-using (aleatory) adviser's influence slightly higher than the confidence term-using (epistemic) adviser (Figure 2.12). The results of the Bayesian ANOVA showed anecdotal evidence regarding the main effect of task ( $BF_{10} = 0.62$ ) and adviser type ( $BF_{10} = 0.36$ ). There was substantial evidence for the interaction term ( $BF_{10} = 4.34$ ).

**Figure 2.12** Adviser Influence by Task in Experiment 2



### 2.3.3 Discussion

The methodological changes resulted in improved experimental manipulations. The adviser distinction score showed that participants were better able to distinguish between the two types of advisers and identify which linguistic expression each adviser used. The Epistemic-Aleatory Rating Scale (EARS) scores were generally as expected and in line with Experiment 1. Participants viewed the historic dates (epistemic) task to have higher levels of epistemic uncertainty compared to the temperature prediction (aleatory) task. Conversely, the perceived level of aleatory uncertainty rating was significantly higher in the temperature prediction (aleatory) task than in the historic dates (epistemic) task. The EARS score differed within each task as well in the expected direction. However, the difference within the temperature prediction (aleatory) task was not significant. Similar to Experiment 1, participants believed that there is some level of epistemic uncertainty associated with future temperature prediction.

The findings on adviser competence rating and adviser influence partially supported the hypotheses. The adviser who gave advice using likelihood terms (aleatory adviser) was rated as more competent in the temperature prediction (aleatory) task compared to the historic dates (epistemic) task. The adviser who used confidence terms (epistemic adviser) was on average perceived as more competent in the historic dates (epistemic) task than in the temperature prediction (aleatory) task, although the difference was not significant. Advisers' influence on participants' answers also differed by task type, where the confidence term-using (epistemic) adviser's influence was greater in the historic dates (epistemic) task compared to the temperature prediction (aleatory) task. The likelihood term-using (aleatory) adviser's influence did not differ by task type. Therefore, the results of adviser competence rating and adviser influence together shows that people are

moderately sensitive to the relationship between advisers' linguistic expression and the type of contextual uncertainty.

These results extend Ülkümen and colleagues' (2016) findings, which showed that people prefer to describe epistemic uncertainty using confidence terms but tend to use likelihood terms to depict aleatory uncertainty. In the current experiment, participants exhibited greater advice adoption for advisers who used context-appropriate linguistic expressions in their advice. This might be because the use of different linguistic expressions signals that the adviser can accurately assess the source of the uncertainty in the advice-taking context, and thus demonstrates greater knowledge and understanding. Participants appear to be sensitive to this subtle signal and infer the adviser's ability from their linguistic expressions. The findings are also in line with the confidence heuristic, which argues that advisees infer adviser accuracy and knowledge from adviser confidence (Price & Stone, 2004). In previous literature, these inferences were made regarding the magnitude of adviser confidence (e.g., numerical value). However, in the current experiments, participants appear to be using advisers' linguistic expressions to construe advisers' knowledge.

On the other hand, in Experiments 1 and 2, participants in each group performed two distinct tasks, and difference in task features may have contributed to the significant results. Hence, a follow-up experiment was conducted where only the participants' perceptions of aleatoriness versus epistemicness was manipulated within a single task. All participants performed a task predicting stock price trends but their perceived level of epistemic and aleatory uncertainty of the stock market was manipulated in Experiment 3.

## 2.4 Experiment 3

Experiment 2 found some evidence that advisees are sensitive to context-appropriate linguistic expressions in advice. Specifically, participants were more likely to adopt advice from an adviser who used context-appropriate linguistic expression, and viewed these advisers as more competent. However, in Experiment 2, two different tasks were used to manipulate the perception of epistemic and aleatory uncertainty. Thus, a follow-up experiment was conducted in which we manipulated the perception of epistemic versus aleatory uncertainty within a single task. Researchers have found that how stock prices are presented (absolute price vs. relative price) could alter the perceived level of epistemic and aleatory uncertainty in the stock market (Walters et al., 2023). Following this method, all participants predicted the future stock price trends, but the information provided to each group of participants varied to manipulate the two types of uncertainty perception.

### 2.4.1 Methods

**Participants.** In total, 60 participants (30 = Female, 30 = Male,  $M_{\text{age}} = 34.6$   $SD_{\text{age}} = 9.7$ ) were recruited through Prolific (<http://prolific.com>). Prior to starting the experiment, all participants provided informed consent in accordance with the University of Oxford Central University Research Ethics Committee guidelines. Participants were compensated £9/hour based on the median duration of the experiment.

**Task.** In Experiment 3, all participants predicted if a firm's stock price would increase or decrease the following month given some firm information from the past 2 years. However, the perceived level of epistemic and aleatory uncertainty of the stock market was manipulated between groups. More specifically, the predictability of the stock trends was highlighted for one group (epistemic group), whereas the volatile features of the stock

market were emphasised for another group (aleatory group). Participants in both groups viewed two advisers: one used confidence terms and the other used likelihood terms.

The stock information used in the experiment was of real firms taken from Yahoo Finance (<https://uk.finance.yahoo.com>). However, the name of the firms was masked, and simply presented as an alphabetical letter (e.g., Stock A). Four pieces of information about the firm were provided to aid decision-making and manipulate participants' perception of the stock market. The first information was this month's stock price. The second was the number of times the stock price increased compared to the previous month in the last 24 months. For example, if in the past 2 years, stock price increased for 16 months, participants were shown "16/24". These two pieces of information were identical for the epistemic and the aleatory versions of the task.

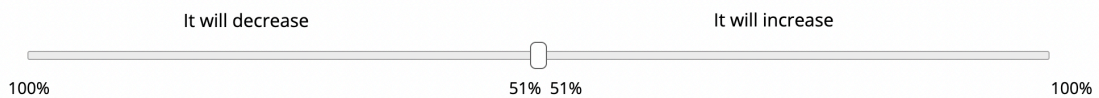
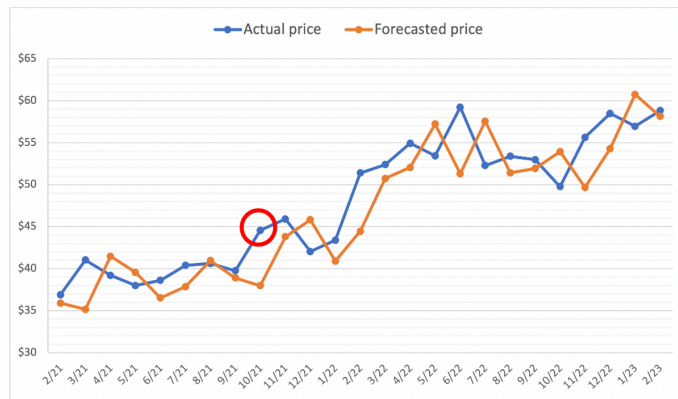
However, the remaining two pieces of information varied by group (epistemic vs. aleatory), which were graphs showing the past stock trend information and a possible reason for change in the stock price from one month to the next. More specifically, for each stock, the monthly stock information for the past 2 years was provided in graph form. In the epistemic group, participants were given the absolute and the forecasted stock price for each month in a line graph. The stock prices showed a distinct increasing or decreasing trend over the past 24 months. The graph was designed to demonstrate that there is a clear trend in stock prices and highlight the similarity between actual and predicted stock prices, which in turn would make the stock market appear more foreseeable. An example is shown in Figure 2.13.

**Figure 2.13** Example question for epistemic group in Experiment 3

Will the average stock price of **stock A** increase or decrease **next month**?

**Firm Information**

- Current (2/23) stock price: \$58.81
- Number of months stock price increased (compared to the previous month) in the last 24 months: 16/24
- Possible reason for change in stock price for circled month: Partnership ended with underperforming company.



In the graph presented to the aleatory group, the stock’s return compared to its previous month were presented for each month. This was calculated using the equation below based on the absolute stock price data presented in the epistemic group. Thus, the data presented to the participants were from the same source but presented in a different format.

$$\text{Stock Return} = \frac{\text{stock price current month} - \text{stock price previous month}}{\text{stock price previous month}}$$

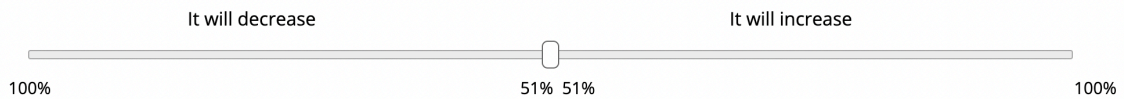
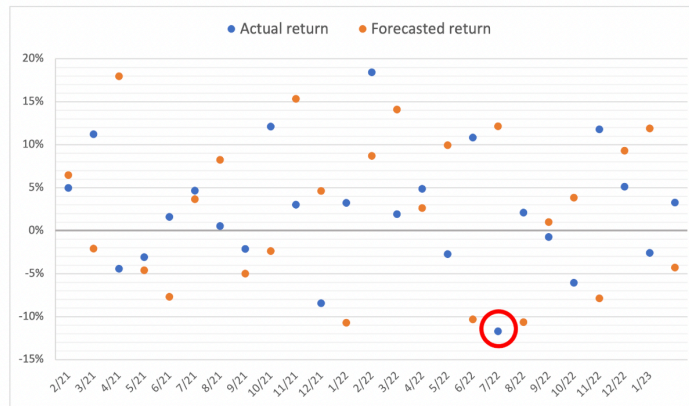
The actual and forecasted stock return for each month was presented as a scatter plot, drawing attention to the variability in the stock prices. An example is shown in Figure 2.14. For both epistemic and aleatory groups, the predicted stock price or return was created by multiplying a random number in the range of 0.95 to 1.05 (i.e., maximum 5% variation) to the actual stock price or return of the previous month. For example, predicted stock price or return in April 2022 was calculated by multiplying a random number (e.g., 1.02) to the actual stock price or return data in March 2022.

**Figure 2.14** Example question for aleatory group in Experiment 3

Will the average stock price of **stock A** increase or decrease **next month**?

**Firm Information**

- Current (2/23) stock price: \$58.81
- Number of months stock price increased (compared to the previous month) in the last 24 months: 16/24
- Possible reason for change in stock price for circled month: New tax policy reduced revenue.



For both groups, one month was circled in red on the graph and a possible explanation for the change in stock price compared to the previous month was provided. For the epistemic group, the reason for change in stock prices were generally given as an internal factor that was somewhat within the firm’s control. For example, the following possible reason was given when the stock price decreased compared to the previous month: **Invested large sum in new product development but was unsuccessful**. If the stock price increased, a reason such as: **New factories reduce production cost** was given. On the other hand, the explanations provided for the aleatory group were generally due to unpredictable or external factors. An example of a reason given for decrease in stock price was, **Industry faces higher environmental tax rate**. An example of an explanation for stock price increase was: **Fall of competitor led to increased market share**.

Based on the four firm information provided, participants were asked to predict if the stock price would increase or decrease the following month. Participants' answers were collected on a scale ranging from 100% confident it will decrease to 100% confident it will increase.

**Adviser.** The adviser's answer and confidence were generated using a similar method as Experiment 1 and 2. There were once again two advisers – one using confidence terms (confident, sure, certain) and one using likelihood terms (chance, probability, likelihood). Adviser's prediction of the stock price or return was generated following the same method as the forecasted price and return but were created independently from the forecast values. More specifically, advisers predicted the next month's stock price or return by multiplying a random number between 0.95 to 1.05 to the current month's stock price. If the random number was less than one, the adviser predicted that the stock price would decrease whereas the adviser predicted an increase when the random number was greater than one. When the random number was exactly one, the advisers randomly predicted an increase or a decrease. Adviser confidence increased as the random number deviated further from one. For example, when the random number was greater than 1.04 or less than 0.96 the adviser's confidence ranged from 91 to 100. Participants once again were not given the adviser's exact prediction but only their confidence and predicted trend (i.e., increase or decrease).

**Procedure.** The experiment used a mixed design in which participants' perceived level of epistemic and aleatory uncertainty of the stock market was manipulated between subjects, but all participants received advice from two types of advisers. Participants were first briefed about the task they will perform, and that they would receive help from

financial advisers. The epistemic group were then told that the stock market is predictable using algorithms and were given example of how a firm's performance influenced their stock price (e.g., Covid-19 vaccine increased the stock price of Pfizer). On the other hand, the aleatory group was told that the stock market is unpredictable and subject to diverse events (e.g., Covid-19 hits the travel industry).

Participants were first given an example question and asked to submit their initial answer. Upon submitting their initial answer, two advisers appeared who used two different linguistic expressions (confidence vs. likelihood terms) but gave the same answer. Participants then submitted their final answer. Next, participants performed six training trials and 18 test trials. The advisers appeared equal number of times and alternated every three trials. The order the advisers were presented was counterbalanced. Lastly, participants reported competence rating for each adviser, performed the adviser distinction questions, and Epistemic Aleatory Rating Scale (EARS) questionnaire (materials available in Appendix A).

**Data Analysis.** Data were analysed following the same method as Experiment 1 and 2. Perceived level of epistemic and aleatory uncertainty in each task was compared through a 2 (task type) x 2 (epistemicness vs. aleatoriness) ANOVA. A t-test comparing the adviser distinction score against three was conducted. Participants' initial accuracy (i.e., accuracy before advice) was compared between the two tasks using a t-test. In order to test the hypotheses, two 2 (task type) x 2 (adviser type) ANOVAs were conducted. One ANOVA examined the two advisers' competence rating and another analysed adviser influence. Adviser influence was calculated based on the change in participants' confidence after advice ( $M\Delta_{\text{confidence}}$  for agreeing trials -  $M\Delta_{\text{confidence}}$  for disagreeing trials).

## 2.4.2 Results

**Table 2.6** *Descriptive statistics of measures in Experiment 3*

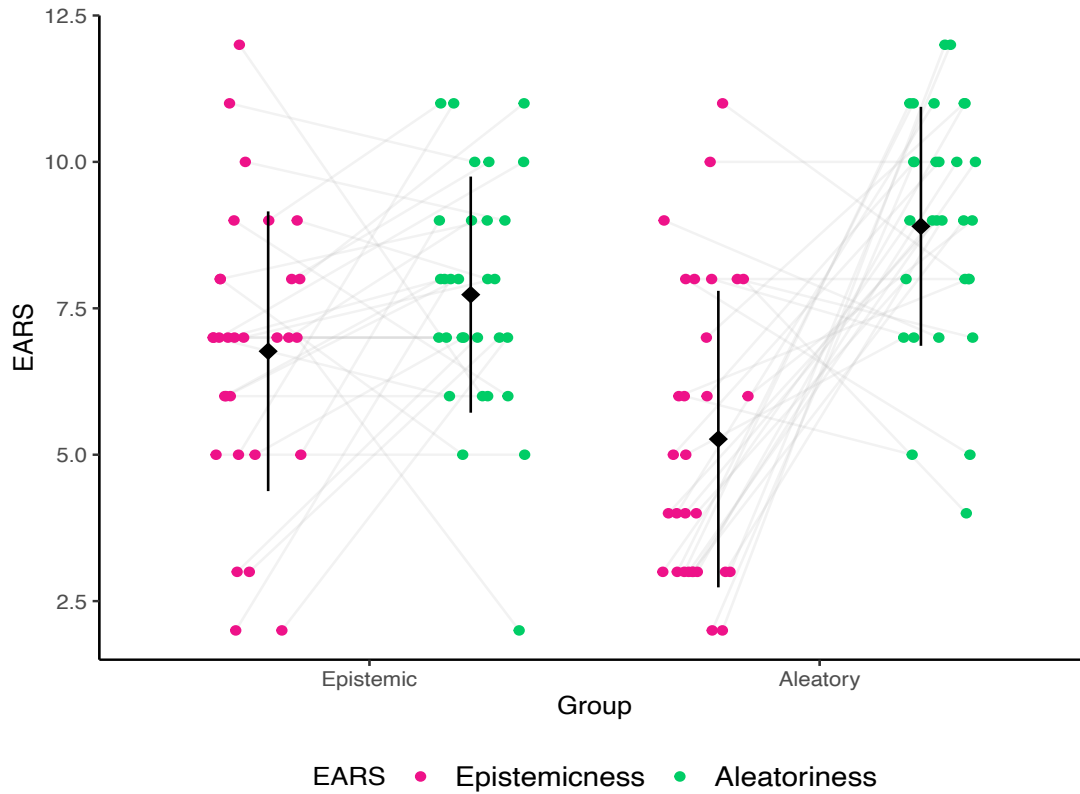
	Epistemic Group		Aleatory Group	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Task Epistemicness	6.8	0.4	5.3	0.5
Task Aleatoriness	7.7	0.4	8.9	0.4
Epistemic Adviser Rating	3.4	0.2	3.2	0.2
Aleatory Adviser Rating	3.4	0.2	3.1	0.2
Epistemic Adviser Influence	18.0	2.2	16.7	2.3
Aleatory Adviser Influence	18.3	3.2	20.1	2.9
Adviser Distinction Score	3.2 (1.4)		4.2 (1.8)	

*Note.* Values in parenthesis are standard deviations.

***Epistemic-Aleatory Rating Scale (EARS).*** The perceived level of epistemic and aleatory uncertainty in each task was collected. A 2 (task type) x 2 (epistemicness vs. aleatoriness) ANOVA was conducted with the rating value as the dependent variable. There was no main effect of task type,  $F < 1$ ,  $\eta^2 = .001$ . However, there was a significant main effect of rating type,  $F(1,58) = 25.50$ ,  $p < .001$ ,  $\eta^2 = .21$ , where the aleatoriness rating was significantly higher than the epistemicness rating. This was driven by the high aleatory uncertainty rating within the epistemic group. The central interest of the analysis was the interaction effect, which was significant,  $F(1,58) = 8.57$ ,  $p = .005$ ,  $\eta^2 = .08$ . Two pairwise comparison with Bonferroni correction were carried out to further examine the interaction effect. The perceived level of epistemic uncertainty was greater in the epistemic group compared to the aleatory group,  $p = .022$ ,  $d = -0.61$ . In contrast, the perceived aleatoriness was higher in the aleatory group compared to the epistemic group,  $p = .030$ ,  $d = 0.58$ . However, within epistemic group, the epistemicness and aleatoriness ratings did not differ significantly,  $p = .096$ ,  $d = 0.44$ , and the numerical trend was towards a higher aleatoriness rating. This was unexpected as task manipulations (e.g., graph and reason for stock price trends) were designed to highlight the epistemic uncertainty of the stock market. On the

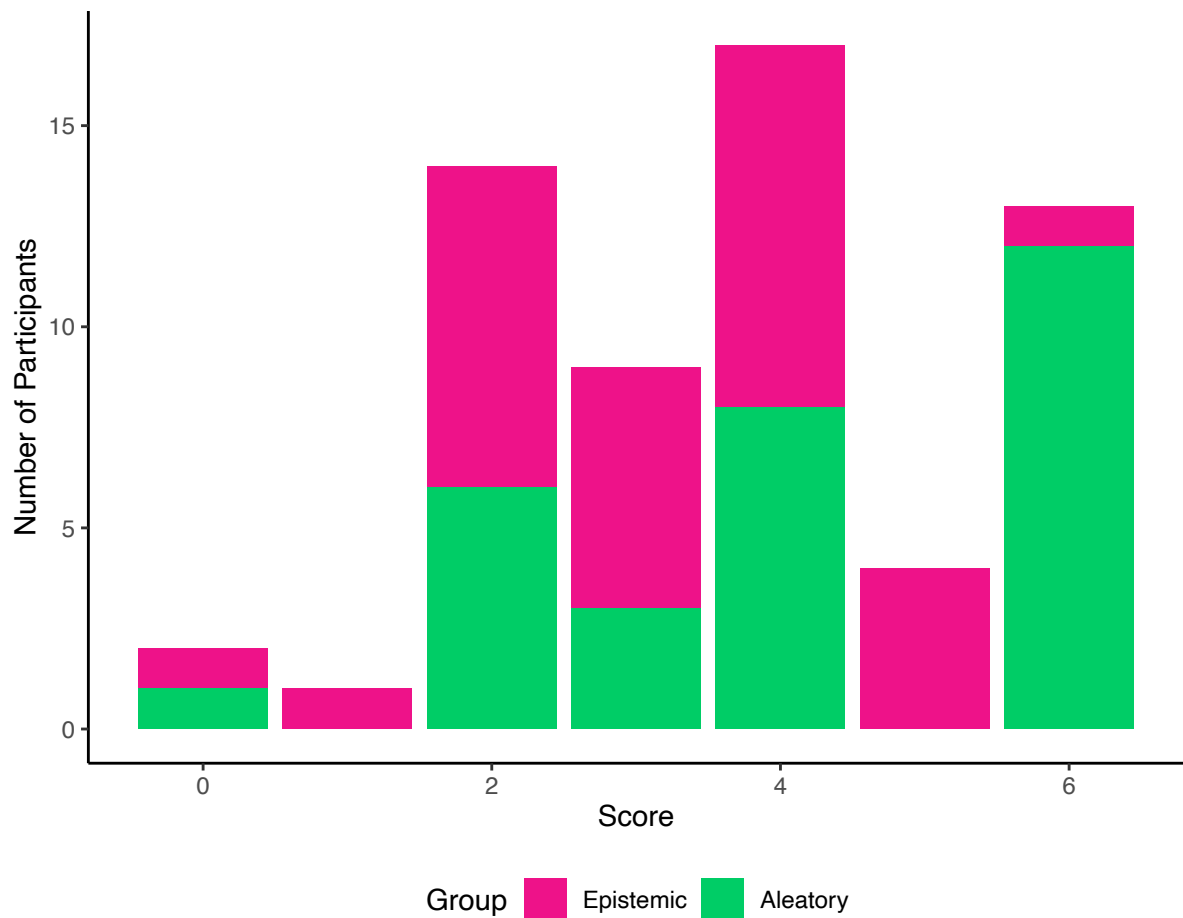
other hand, as expected, the perceived aleatory uncertainty was significantly higher than the perceived epistemic uncertainty within the aleatory group,  $p < .001$ ,  $d = 1.58$ .

**Figure 2.15** *Epistemic-Aleatory Rating Scale (EARS) by task in Experiment 3*



**Adviser Distinction.** Adviser distinction questions were collected to test if participants were aware that the two advisers were using different linguistic expressions. The overall adviser distinction score was significantly different from three ( $M = 3.7$ ,  $SD = 1.63$ ),  $t(59) = 3.33$ ,  $p = .002$ ,  $d = 0.43$ . Additionally, most participants scored higher than three (Figure 2.16). This shows that participants were able to identify the statements used by each adviser above chance level.

**Figure 2.16** Adviser distinction score by task in Experiment 3



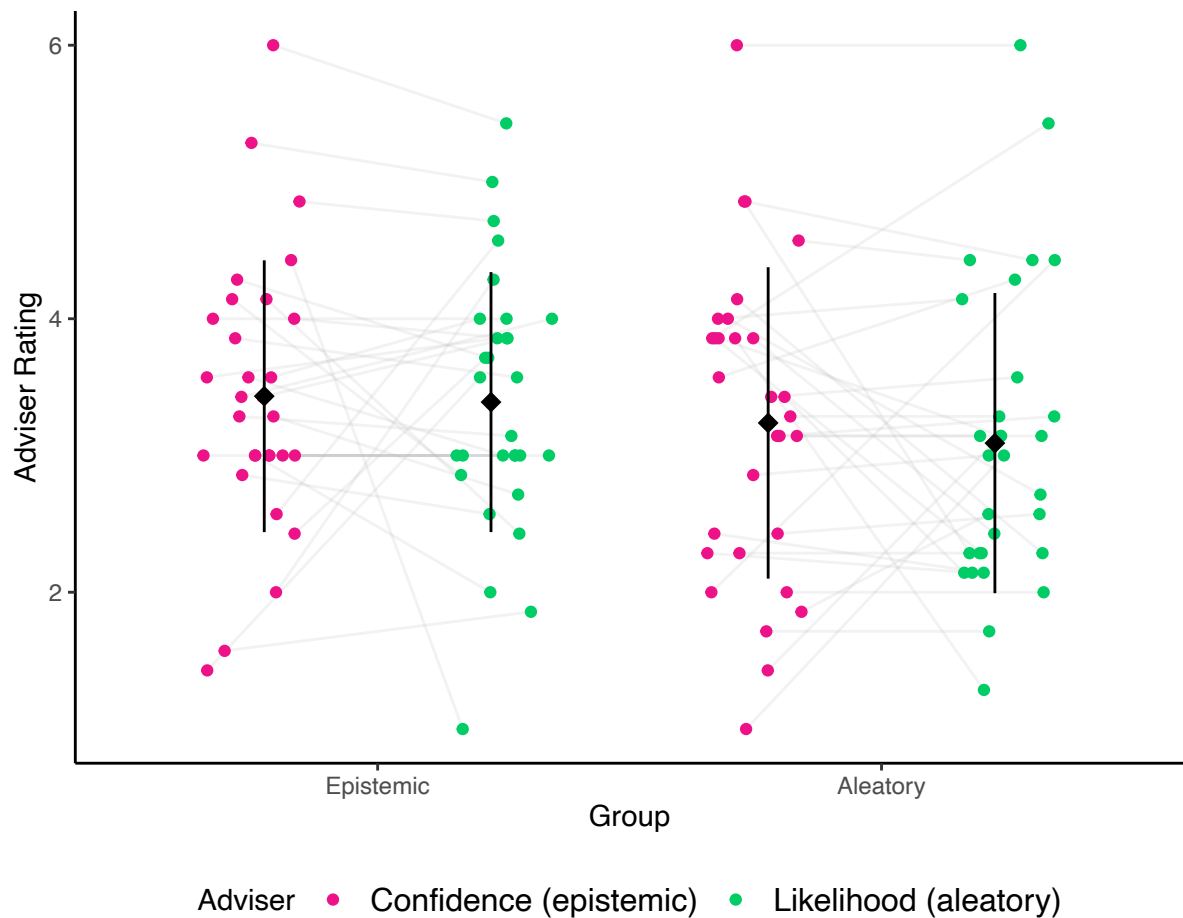
**Accuracy.** Participants' accuracy was calculated based on the actual stock price on March 2023 (experiment was conducted in February 2023), which was collected from Yahoo Finance (<https://uk.finance.yahoo.com>). The accuracy of the participants who viewed the stock price (epistemic group) and participants who were shown the stock return (aleatory group) did not significantly differ,  $t < 1$ ,  $d = -0.11$ . A 2 (initial vs. final) x 2 (epistemic vs. aleatory) ANOVA found no significant main effect of time point,  $F < 1$ ,  $p = .909$ ,  $\eta^2 < .001$ . These contradicted results of Experiments 1 and 2 which observed significant improvement in accuracy after advice adoption. The no change in accuracy may be due to the randomness in the stock market. The main effect of task type,  $F < 1$ ,  $p = .560$ ,  $\eta^2 = .002$ , and the interaction effect,  $F < 1$ ,  $p = .560$ ,  $\eta^2 = .002$ , was not significant.

**Table 2.7** Participant accuracy by task in Experiment 3

	Epistemic group		Aleatory group	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Initial Accuracy (%)	45	0.10	46	0.11
Final Accuracy (%)	47	0.11	46	0.12

**Adviser Rating.** The EARS score and adviser distinction score demonstrated that the experimental manipulations were mostly successful. The main analysis of interest was whether advisers' use of different linguistic expressions in their advice had an effect on perceived adviser competence (H1 & H2). Adviser competency rating was analysed through a 2 (task type) x 2 (adviser type) ANOVA with adviser rating as the dependent variable. There was no significant main effect of task type,  $F(1, 58) = 1.196, p = .279, \eta^2 = .01$ , or adviser type,  $F < 1, \eta^2 = .002$ . The main effects were not expected to be significant and the main interest of the analysis of the interaction effect. However, the interaction effect was also not significant,  $F < 1, \eta^2 < .001$ . The numerical difference in adviser rating was also minimal as visible in Figure 2.17. The results of the Bayesian ANOVA showed similar patterns. Analysis found anecdotal evidence against the main effect of task ( $BF_{10} = 0.45$ ). The evidence against the main effect of adviser type was substantial ( $BF_{10} = 0.22$ ). There was also substantial evidence against the interaction term ( $BF_{10} = 0.28$ ). Therefore, no evidence in support of the hypotheses were found.

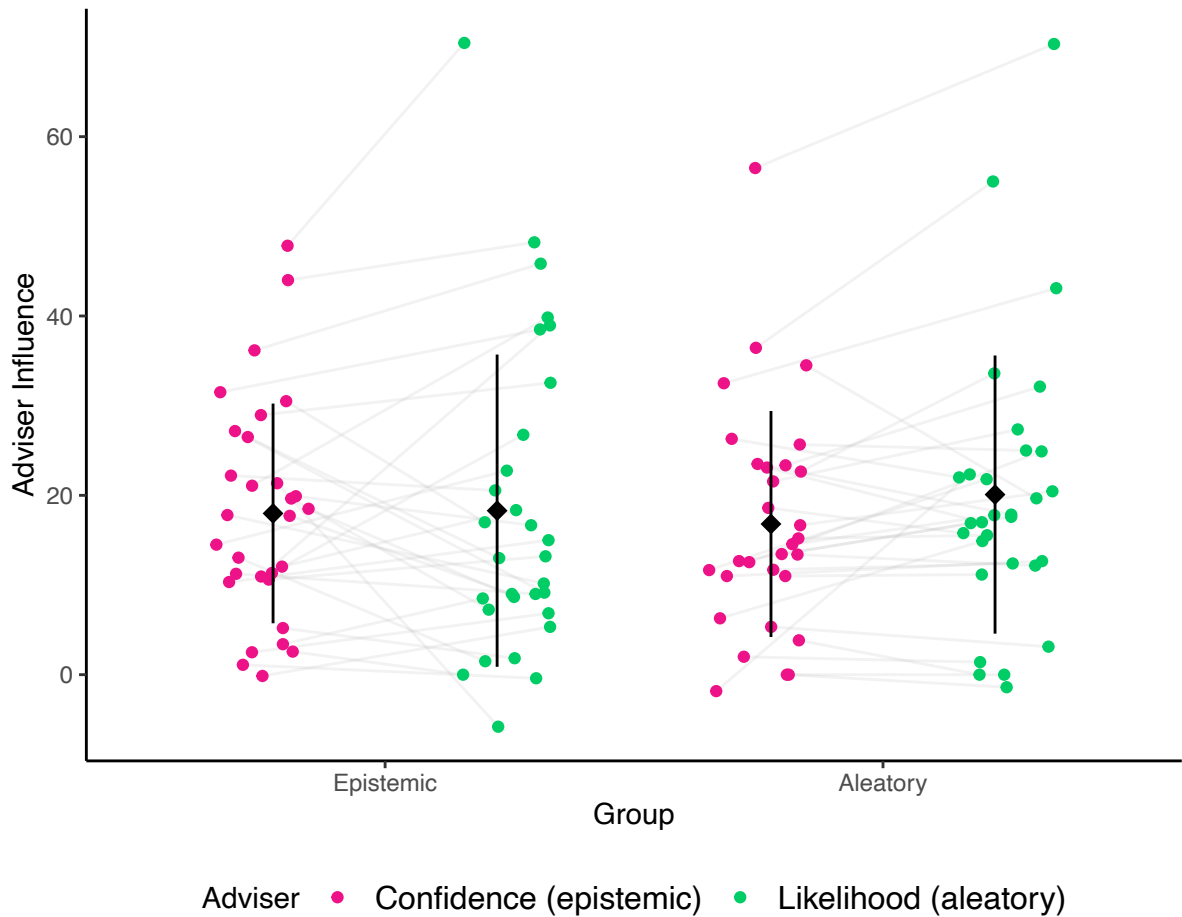
**Figure 2.17** *Adviser competency rating by task in Experiment 3*



**Adviser Influence.** The second analysis of interest was the effect of advisers' different linguistic expressions on advice adoption (H1 & H2). Once again, a 2 (task type) x 2 (adviser type) ANOVA was conducted. There was no significant main effect of task type,  $F < 1, \eta^2 < .001$ . The main effect of adviser type was also not significant,  $F(1, 58) = 1.83, p = .181, \eta^2 = .004$ . Of critical interest was the interaction effect which was not significant,  $F(1, 58) = 1.26, p = .266, \eta^2 = .003$ . Therefore, no evidence was found to support the hypotheses. However, within the aleatory group, the adviser using likelihood terms was numerically more influential compared to the adviser using confidence terms (Figure 2.18). Bayesian ANOVA was conducted to better understand the lack of significant results compared to Experiment 2. The analysis showed anecdotal evidence against the

main effects (task  $BF_{10} = 0.41$ ; adviser  $BF_{10} = 0.43$ ). There was also anecdotal evidence for the absence of an interaction effect ( $BF_{10} = 0.46$ ).

**Figure 2.18** *Adviser influence by task in Experiment 3*



### 2.4.3 Discussion

In Experiment 3, participants were asked to predict if a stock price would increase or decrease the following month. The task manipulation was mostly successful where the perceived epistemic uncertainty was higher in the epistemic group compared to the aleatory group. The aleatory group was also perceived to have higher levels of aleatory uncertainty compared to the epistemic group. Additionally, participants believed the aleatory group to have higher levels of aleatory uncertainty compared to epistemic uncertainty. These findings were consistent with Walters and colleagues (2023) results, which showed that presenting absolute stock price increases perceived epistemic

uncertainty of the stock market whereas viewing relative stock prices makes the stock market appear high in aleatory uncertainty. However, within the epistemic group, the perceived aleatory uncertainty rating was numerically greater than the epistemic uncertainty rating. This may be because people inherently perceive the stock market as unpredictable and thus, the task manipulations were insufficient to change this belief.

However, despite the moderately successful experimental manipulations, minimal effects of advisers' varying linguistic expressions were found. There were minor differences in adviser rating and adviser influence across and within the two tasks. The only notable difference was that the adviser using likelihood terms (aleatory adviser) was slightly more influential than the adviser using confidence terms (epistemic adviser) within the aleatory group. However, the difference was not significant and comparison within the task was not the main objective of the experiment. One possible reason for lack of findings was that the perceived aleatory uncertainty was numerically higher than perceived epistemic uncertainty in the epistemic group. It is possible that the task manipulation needs to be more distinct for the effects to be detected. Therefore, another experiment was conducted with the task manipulation as the main focus.

## **2.5 Experiment 4**

The lack of findings in Experiment 3 was puzzling considering the significant results in Experiment 2. In order to better understand the discrepancy in the findings, another experiment was conducted. In Experiment 4, tasks with greater disparity in epistemic and aleatory uncertainty were created. In one task participants were asked to indicate which country has a greater total area, which was designed to have high levels of epistemic

uncertainty but low levels of aleatory uncertainty. The task designed to have high levels of aleatory uncertainty was to predict the outcome of a given number of coin tosses.

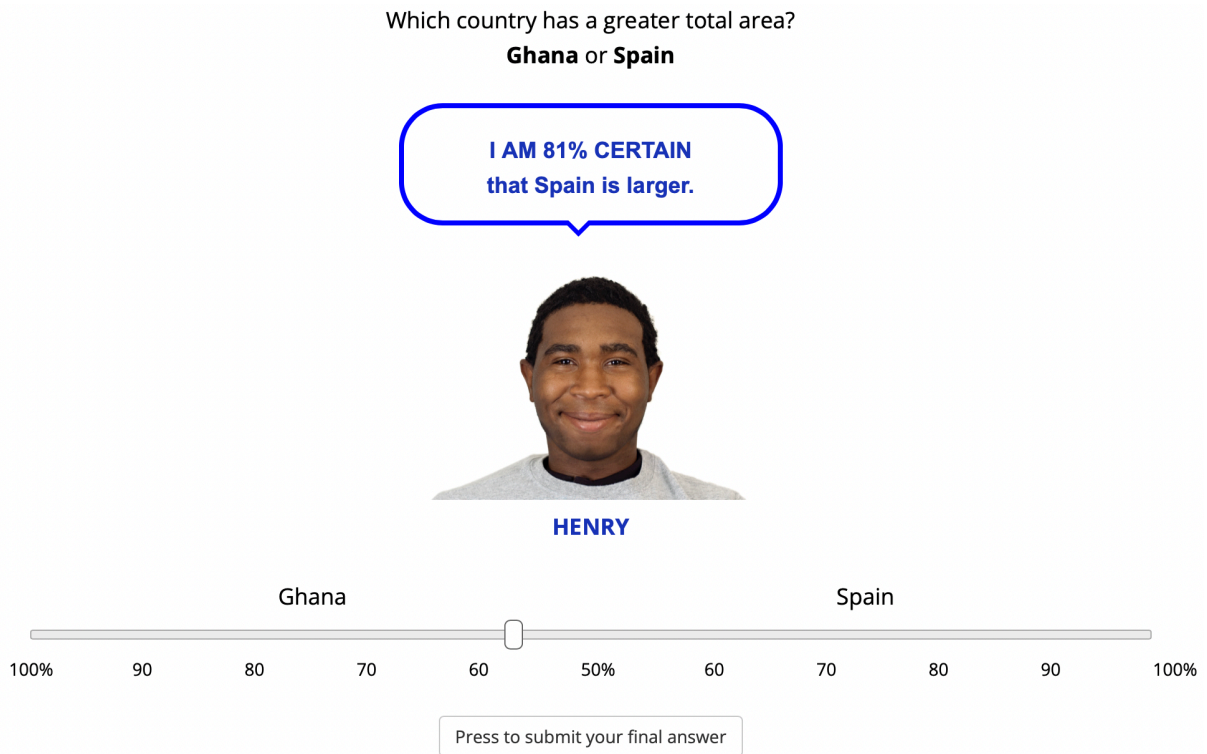
### 2.5.1 Methods

**Participants.** Sixty participants (28 = Female, 30 = Male, 1 = Non-binary, 1 = Prefer not to say,  $M_{\text{age}} = 35.3$ ,  $SD_{\text{age}} = 10.1$ ) were recruited through Prolific (<http://prolific.com>). All participants gave informed consent before commencing the experiment. The experiment received approval from the University of Oxford Central University Research Ethics Committee. Compensation was £9/hour based on the median duration of the experiment.

**Tasks.** There were two tasks, which were designed so that one would be high in epistemic uncertainty, and one would have high levels of aleatory uncertainty. The overall set-up of both tasks was similar to previous experiments, which adopted the judge-adviser system (Sniezek & Buckley, 1995).

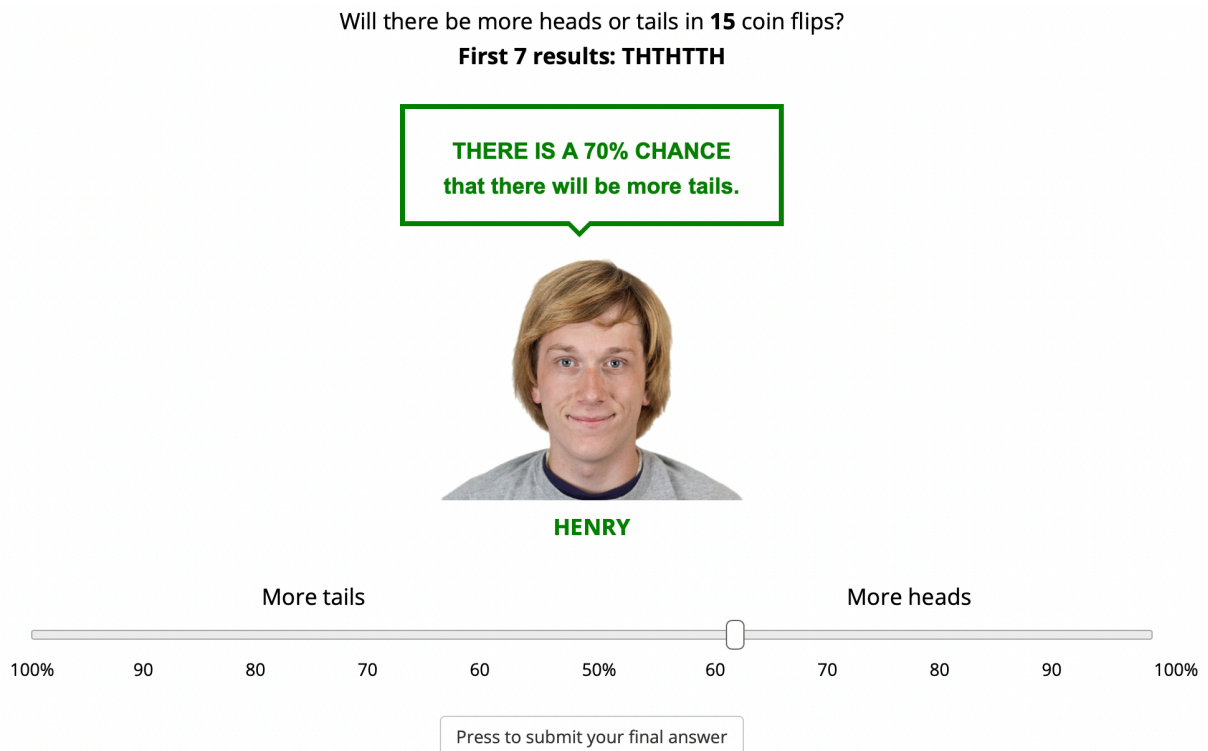
In the task designed to have high levels of epistemic uncertainty (i.e., epistemic task), participants were given two countries and were asked to guess which country has a greater total area. As visible in Figure 2.19, answers were collected on a scale ranging from 100% confident country A to 100% confident country B. The slider started in the midpoint (50%) of the scale and participants had to provide an answer to proceed. After submitting their initial answer, an adviser appeared and gave their answer and accompanying confidence level. Participants were once again asked to provide their answer (final answer). Here, the slider starting point was participants' initial answer.

**Figure 2.19** Example question of the total area (epistemic) task in Experiment 4



In the task designed to have high levels of aleatory uncertainty, participants predicted if there would be more heads or tails in a given number of coin tosses. More specifically, participants were given the first few results of a coin toss (e.g., first five results) and then guessed if there would be more heads or tails overall in a given number (e.g., 13) of coin tosses. The midpoint was once again unselectable to force an answer. Participants' answer was collected pre- and post-advice. Participants gave their answer on a scale ranging from 100% heads to 100% tails. An example of the task is displayed in Figure 2.20.

**Figure 2.20** Example question of the coin toss (aleatory) task in Experiment 4



*Adviser.* In Experiment 3, participants distinguished the two advisers above chance level, but the adviser distinction score was still low ( $< 4$ ). Therefore, in Experiment 4, a new picture from the Chicago Face Database (<https://www.chicagofaces.org>) was selected to represent the two advisers. The advisers were both male, but one adviser was Black, and the other adviser was White. This was to aid distinction of the two advisers. The advisers were of similar levels of attractiveness and trustworthiness. Additionally, the two advisers' speech bubble varied in terms of shape (square vs. round edges) and colour (green vs. blue). The picture representing each adviser, the shape of the speech bubbles, and the colour of the texts were all randomised.

In the total area (epistemic) task, adviser confidence was determined by the magnitude of the total area difference between the two countries. If the two countries had similar total areas, the adviser displayed low confidence whereas a greater percentage difference in

the total area resulted in higher confidence. For example, Romania has a total area of 238,397 square kilometres whereas Cambodia's total area is 181,035 square kilometres. Therefore, Romania is approximately 32% larger than Cambodia. In this case the adviser's confidence was set to 62%. On the other hand, the total area percentage difference between Ghana and Spain (Figure 2.19) is 112%. Here, the adviser was 77% confident. One major difference from previous experiments was that an exact confidence value was first calculated based on the total area percentage difference and then a random number between -10 and +10, generated from a truncated normal distribution ( $M = 0, SD = 8$ ), was added to the confidence value. For example, in Figure 2.19, the adviser confidence was initially set to 77% and then 4 (random number) was added, and thus final adviser confidence was 81%. A random number was added to weaken the correlation between actual total area and adviser confidence. This method was used instead of selecting a random number between a range based on the total area difference (e.g., adviser's confidence is between 51% to 60% when total area difference is below 30%) due to the high variability in the total area difference. In the two examples provided above, there is an 80% difference between the total area percentage (32% vs. 112%). Therefore, it was difficult to set an appropriate confidence range based on the total area percentage difference. Additionally, we wished to use the same method as the coin toss (aleatory) task to calculate adviser confidence.

In the coin toss (aleatory) task, adviser confidence was calculated using the binomial probability formula based on the current coin toss results and the remaining number of coin tosses. For example, suppose that a total of seven consecutive coin tosses are to be performed and results of the first three tosses were heads, heads, tails. For there to be more heads overall, there needs to be at least two heads in the remaining four coin tosses,

for which there is a 68.75% probability. Therefore, adviser confidence was calculated by first rounding the probability to the nearest integer (e.g., 69%) and adding a random number between -10 and +10 (from normal distribution). In both tasks, the advisers' confidence was bounded to be between 51% to 100%.

***Procedure.*** The experiment used a mixed design where participants either performed the total area (epistemic) or coin toss (aleatory) task. However, in both tasks, there were two advisers with one using confidence terms and another using likelihood terms.

All participants first gave informed consent before starting the experiment. After reading task instructions, participants completed the training block consisted of six trials (three per adviser). Participants were given a chance to take a break then engaged in the test block which had 18 trials. Lastly, participants completed the adviser competency questionnaire for each adviser, six adviser distinction questions, and the Epistemic-Aleatory Rating Scale (EARS). Upon completion, participants were debriefed.

***Data Analysis.*** Data analysis method was identical to the previous experiments. The perceived level of epistemic and aleatory uncertainty was compared between the two tasks using a 2 (task type) x 2 (epistemicness vs. aleatoriness) ANOVA. The adviser distinction score was tested against three using a t-test. Participants' initial accuracy was compared between the two tasks through a t-test. Perceived adviser competence and adviser influence was compared using two separate 2 (task type) x 2 (adviser type) ANOVAs.

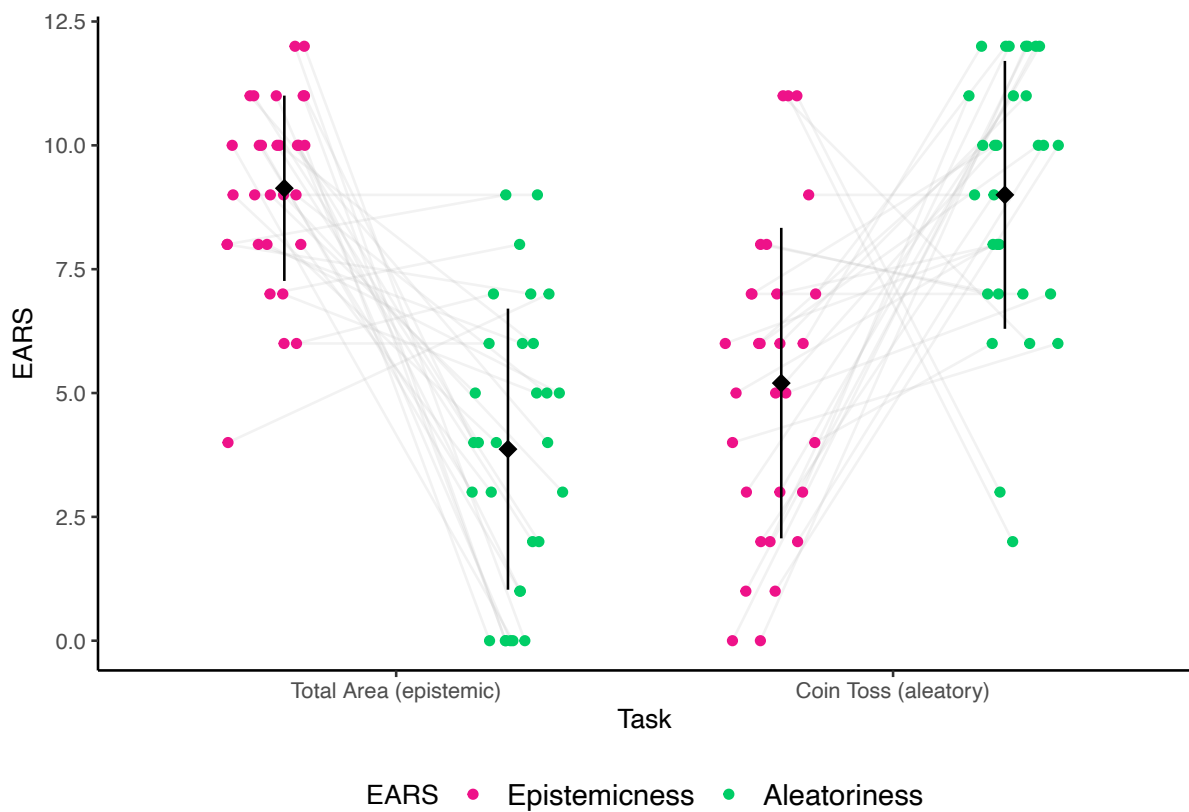
## 2.5.2 Results

**Table 2.8** *Descriptive statistics of measures in Experiment 4*

	Total Area (epistemic) task		Coin Toss (aleatory) task	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Task Epistemicness	9.1	0.3	5.2	0.6
Task Aleatoriness	3.9	0.5	9	0.5
Epistemic Adviser Rating	3.4	0.2	2.7	0.2
Aleatory Adviser Rating	3.6	0.2	3.0	0.2
Epistemic Adviser Influence	28.7	4.6	1.6	0.7
Aleatory Adviser Influence	27.7	4.4	4.1	1.6
Adviser Distinction Score	2.9 (2.0)		3.7(2.3)	

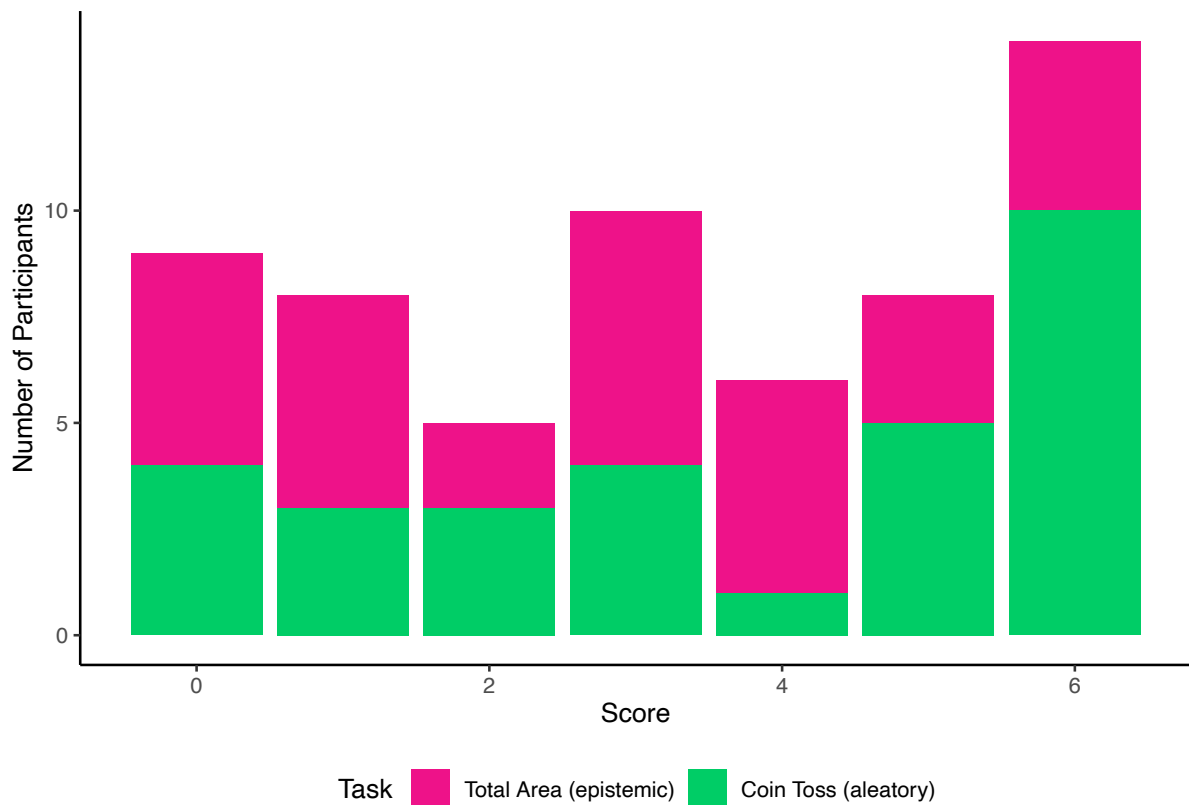
**Epistemic-Aleatory Rating Scale (EARS).** The EARS was collected to examine if perceived level of epistemic and aleatory uncertainty differed by task type. A 2 (task type) x 2 (epistemicness vs. aleatoriness) ANOVA revealed that there is a significant main effect of task type,  $F(1,58) = 4.43, p = .040, \eta^2 = .01$ . The overall rating was higher in the coin toss (aleatory) task, where participants perceived the task to have some levels of epistemic uncertainty. However, the main effect of rating type was not significant,  $F(1,58) = 1.36, p = .249, \eta^2 = .02$ . Of critical interest was the interaction effect between task type and rating type, which was significant,  $F(1,58) = 51.80, p < .001, \eta^2 = .43$ . Therefore, pairwise comparison with Bonferroni correction was carried out. This showed that the epistemicness rating was significantly higher in the total area task compared to the coin toss task,  $p < .001, d = -1.52$ . Similarly, the aleatoriness rating was significantly higher in the coin toss task,  $p < .001, d = 1.85$ . The rating within each task also differed significantly. In the total area (epistemic) task, the epistemicness rating was significantly greater than the aleatoriness rating,  $p < .001, d = -2.19$ . The opposite was true for the coin toss (aleatory) task,  $p < .001, d = 1.30$ . Therefore, manipulation of the tasks to have different levels of epistemic and aleatory uncertainty was successful.

**Figure 2.21** Epistemic-Aleatory Rating Scale (EARS) by task in Experiment 4



**Adviser distinction.** Participants were asked to pair six statements (involving three confidence terms and three likelihood terms) with each adviser to examine if participants can distinguish between the two advisers. Surprisingly, the adviser distinction score was not significantly different from three (chance level) ( $M = 3.27$ ,  $SD = 2.17$ ),  $t < 1$ ,  $d = 0.12$ . There were some participants who scored six, but most participants scored below three (Figure 2.22). This was unexpected as participants' adviser distinction score was above chance level in Experiment 2 and 3. Higher adviser distinction scores were expected due to the further efforts made in the current experiment to aid distinction (e.g., adviser pictures, colours of speech bubble).

**Figure 2.22** Adviser distinction score by task in Experiment 4



**Accuracy.** In the total area (epistemic) task, participants accuracy was calculated based on the correct answer (i.e., picked country with larger total area). On the other hand, in the coin toss (aleatory) task, the probabilistically more likely outcome between heads and tails was considered as the correct answer and accuracy was calculated accordingly. A Welch’s two-sample t-test was conducted due to the unequal variance in accuracy between the two tasks. The analysis found that initial accuracy of participants performing the coin toss (aleatory) task was significantly higher than participants in the total area (epistemic) task,  $t(48.5) = -4.64, p < .001, d = 1.00$ . This was because most participants were accurate 90% of the trials in the coin toss (aleatory) task, whereas participants’ accuracy were about 63% on average in the total area (epistemic) task (Table 2.7). This was also reflected in the ANOVA with task type (total area vs. coin toss) and time point (initial vs. final) as the independent variables and accuracy as the dependent variable. The

analysis revealed a significant main effect of task type,  $F(1, 58) = 10.18, p = .002, \eta^2 = .14$ , demonstrating that accuracy was higher in the coin toss task. The main effect of time point was also significant  $F(1, 58) = 43.19, p < .001, \eta^2 = .07$ . The interaction between task type and time point was also significant,  $F(1, 58) = 30.53, p < .001, \eta^2 = .05$ . Pairwise comparison with Bonferroni correction found that participants' initial accuracy significantly differed by task type ( $p < .001, d = 1.2$ ) but difference in final accuracy was not significant ( $p = .771, d = 0.3$ ).

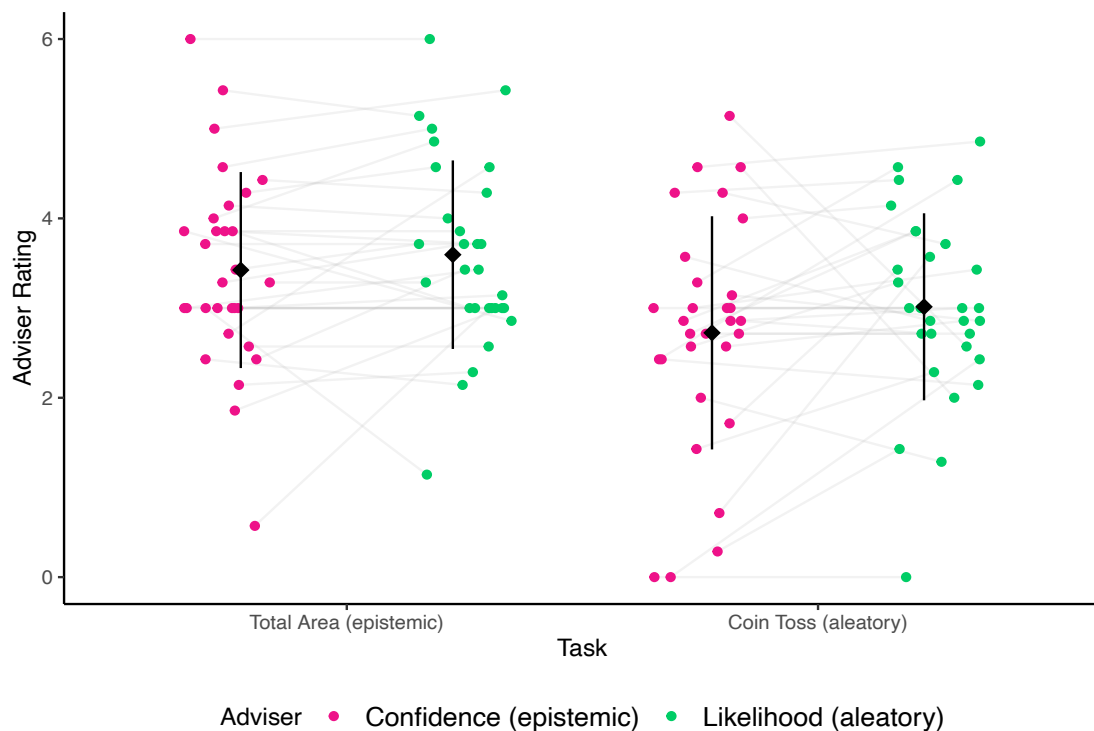
**Table 2.9** Participant accuracy by task in Experiment 4

	Total Area (Epistemic) Task		Coin Toss (Aleatory) Task	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Initial Accuracy (%)	63	0.15	87	0.24
Final Accuracy (%)	82	0.16	89	0.20

**Adviser Rating.** One of the main analyses of interest was to examine if advisers' linguistic expressions influence perceived adviser competence (H1 & H2). A 2 (task type) x 2 (adviser type) ANOVA was conducted to examine if there was a difference in adviser rating. There was a significant main effect of task type,  $F(1, 58) = 6.04, p = .017, \eta^2 = .08$ , where the adviser rating was greater in the total area (epistemic) task compared to the coin toss (aleatory) task. This is likely due to the inherent randomness in the coin toss (aleatory) task, which reduced reliance on the advisers. The main effect of adviser type was not significant,  $F(1, 58) = 3.20, p = .079, \eta^2 = .01$ . The main interest of the analysis was the interaction effect, which was also not significant,  $F < 1, \eta^2 < .001$ . Due to the lower overall adviser rating in the coin toss (aleatory) task, it was difficult to compare advisers' rating across tasks. However, it was expected that there would be a difference within each task. Within the coin toss (aleatory) task the likelihood term-using (aleatory) adviser's rating was numerically higher (Figure 2.23). However, this was also the case

within the total area (epistemic) task, when the opposite trend was expected. A Bayesian ANOVA was conducted to better understand the unexpected results. The analysis found substantial evidence in favour of the main effect of task type ( $BF_{10} = 3.33$ ), confirming that adviser rating was higher in the total area (epistemic) task. The evidence against the main effect of adviser type was anecdotal ( $BF_{10} = 0.80$ ). The main interest was the Bayes Factor of the interaction term which was obtained by comparing the main effects only model and full (main effects and interaction term). There was substantial evidence against the interaction term ( $BF_{10} = 0.27$ ). Therefore, no support for the interaction term was found.

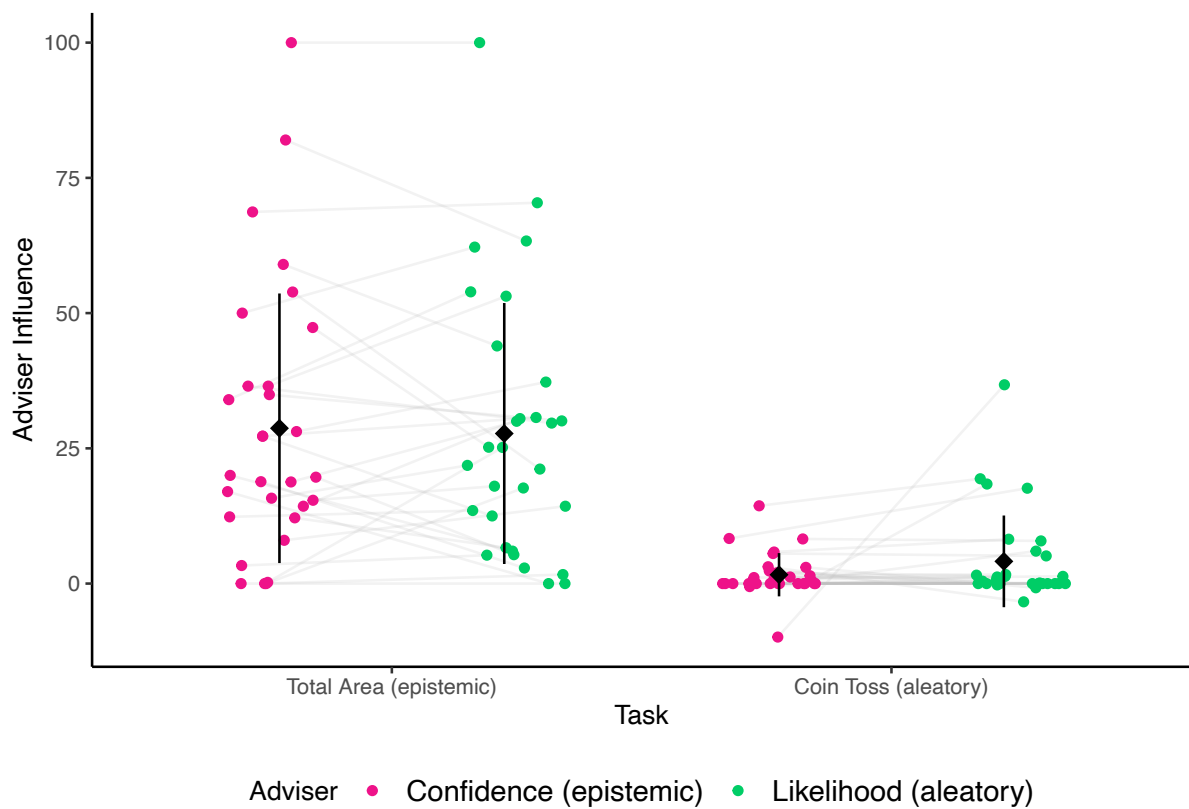
**Figure 2.23** *Adviser competency rating by task in Experiment 4*



**Adviser Influence.** The second main analysis of interest was the impact of advisers' varying linguistic expressions on advice adoption (H1 & H2). Adviser influence was measured as the change in participants' confidence following advice. A 2 (task type) x 2 (adviser type) ANOVA showed that there was a significant main effect of task type,  $F(1, 58) = 33.59, p < .001, \eta^2 = .34$ . The lower overall adviser influence in the coin toss

(aleatory) task is once again presumably due to the inherent randomness in the task, which reduced the perceived utility of advice. The main effect of adviser type was not significant,  $F < 1$ ,  $\eta^2 < .001$ . Critically, the interaction effect was not significant,  $F(1, 58) = 1.25$ ,  $p < .269$ ,  $\eta^2 = .002$ . Therefore, no support for the hypotheses was found. Additionally, no significant differences were observed within each task. However, adviser influence of the likelihood term-using adviser was numerically higher than the confidence term-using adviser within the coin toss (aleatory) task (Figure 2.24). This pattern was also observed in a Bayesian ANOVA, which showed decisive evidence in favour of the main effect of task type ( $BF_{10} = 30343.56$ ). There was substantial evidence against the main effect of adviser type ( $BF_{10} = 0.21$ ). Critically, there was anecdotal evidence against the interaction term ( $BF_{10} = 0.44$ ).

**Figure 2.24** Adviser influence by task in Experiment 4



### 2.5.3 Discussion

The lack of findings in Experiment 3 compared to Experiment 2 were possibly attributed to the limitations in task uncertainty manipulation. Therefore, Experiment 4 was conducted to better understand the inconsistency in the results. In particular, two tasks were used - each designed to have extremely high levels of epistemic or aleatory uncertainty.

The task epistemic and aleatory uncertainty manipulation was successful. The perceived epistemic uncertainty was higher in the total area (epistemic) task compared to the coin toss (aleatory) task. Participants also rated the coin toss (aleatory) task to have higher levels of aleatory uncertainty compared to the total area (epistemic) task. Furthermore, perceived epistemic and aleatory uncertainty ratings also significantly differed within each task in the expected direction. Therefore, the variations in the level of epistemic and aleatory uncertainty in the experimental tasks were the most evident in Experiment 4.

In light of the successful task manipulation, difference in the adviser using confidence terms (epistemic adviser) and the adviser using likelihood terms (aleatory adviser) was expected across the two tasks. However, there were no evidence found to support the hypotheses. In addition to hypotheses testing, adviser rating and influence within each task was also examined. Within the coin toss (aleatory) task, the adviser using likelihood term (aleatory adviser) was slightly more influential than the confidence term-using (epistemic) adviser, but the difference was not significant. There were no other within task differences of adviser rating and influence.

Although this was not the main focus of the experiment, one observed pattern was that adviser rating and influence were significantly lower in the coin toss (aleatory) task compared to the total area (epistemic) task, which hindered comparison across tasks. The high level of perceived aleatory uncertainty in the coin toss (aleatory) tasks might have decreased the perceived utility of advice-taking. This pattern has also been observed in previous literature, which found that increased perceived aleatory uncertainty in the stock market led to decrease in financial advice-taking (Walters et al., 2023). This tendency to disregard advice in tasks with higher aleatory uncertainty was a concern in previous experiments. Therefore, only tasks with moderate levels of aleatory uncertainty were used (e.g., temperature prediction). Despite this risk, a highly random (i.e., coin toss) task was selected for Experiment 4, as it was expected that greater difference in task manipulation would help better detect the effects of adviser's varying linguistic expressions.

One possible explanation for the lack of findings in the current Experiment might be because participants had difficulties distinguishing between the two advisers. In Experiment 4, participants' adviser distinction score was not significantly better than chance level. This was unexpected as extra measures were taken to facilitate the distinction between the two advisers (e.g., new adviser pictures, colour of text) compared to Experiment 2 and 3. Yet, participants were worse at distinguishing between the two advisers in Experiment 4. The difficulties in distinguishing between the two advisers is unlikely to account for all results in the current experiment but might have contributed.

Despite the methodological shortcoming, no difference in adviser rating and influence between the two advisers is still puzzling considering the significant findings in Experiment 2. Additionally, the high aleatory uncertainty in the coin toss task reduced

the overall perceived competence and influence of advisers, which made it difficult to compare the advisers across the two tasks. In order to clarify the findings, an exact replication of Experiment 2 was conducted.

## **2.6 Experiment 5**

Experiment 4 found no evidence to suggest that advisees are sensitive to the relationship between advisers' linguistic expressions and the type of contextual uncertainty prevalent in the advice-taking context. The methodological limitations of Experiment 4 (e.g., difficulties distinguishing between the two advisers, extreme aleatory uncertainty) impeded hypotheses testing. Additionally, the inconsistency in findings among Experiment 2-4 needs clarification. Therefore, an exact replication of Experiment 2 was conducted to test if similar results would be obtained under the same experimental conditions.

### **2.6.1 Methods**

**Participants.** Sixty participants (31 = Female, 29 = Male,  $M_{\text{age}} = 34.2$   $SD_{\text{age}} = 12.1$ ) were recruited through Prolific (<http://prolific.com>). All participants gave informed consent in accordance with the University of Oxford Central University Research Ethics Committee guidelines. Participants were compensated £8/hour based on the median duration of the experiment.

**Tasks.** The tasks used in Experiment 5 were identical to those in Experiment 1 and 2. One task was to estimate if a historic event occurred before or after an anchor year (i.e., epistemic task). Participants' answers and associated confidence were collected on a scale ranging from 100% before to 100% after. The second task was to predict if the

temperature of a given city on a specific future date would be higher or lower than an anchor temperature (i.e., aleatory task). Answers were collected on a scale ranging from 100% lower to 100% higher. In both tasks, participants' initial answer was first collected. After, advice was given and participants' final answer was recorded (i.e., judge-adviser-system)

**Adviser.** Same pictures from Experiment 2 were used to represent the advisers, which were retrieved from the Chicago Face Database (<https://www.chicagofaces.org>). The advisers were given names (James and Henry), and the shape of the speech bubble differed (square vs. round edges). One adviser used confidence (confident, sure, certain) terms and the other used likelihood (chance, probability, likelihood) terms in their advice. Advice and accompanying adviser confidence was generated following the same method as Experiment 1. Which adviser was presented first, and the picture and speech bubble assigned to each adviser were all randomised.

**Procedure.** The overall procedure was identical to Experiment 1 and 2. Participants only performed one task – either guessed year of historic events (epistemic task) or predicted future temperature (aleatory task). In each task, participants interacted with both advisers (using confidence vs. likelihood terms). Participants performed a training block comprised of six trials and then engaged in the test block consisted of 18 trials. After, participants rated the perceived competence of each adviser, completed the six adviser distinction questions and the Epistemic-Aleatory Rating Scale (EARS). Participants were then debriefed.

**Data Analysis.** Data analysis followed the same method as previous experiments. Participants' perceptions of epistemic and aleatory uncertainty in each task were compared using a 2 (task type) x 2 (epistemicness vs. aleatoriness) ANOVA. The adviser distinction score was tested against three using a t-test to check if participants could distinguish between the two advisers above chance level. Participants' initial accuracy in the historic dates (epistemic) task and temperature prediction (aleatory) task was compared through a t-test. Next, in order to test the main hypotheses, perceived competence of the two advisers (confidence term vs. likelihood term) was compared across the two tasks using a 2 (task type) x 2 (adviser type) ANOVA. The adviser's influence on participants' answer was also tested using a 2 (task type) x 2 (adviser type) ANOVA.

### 2.6.2 Results

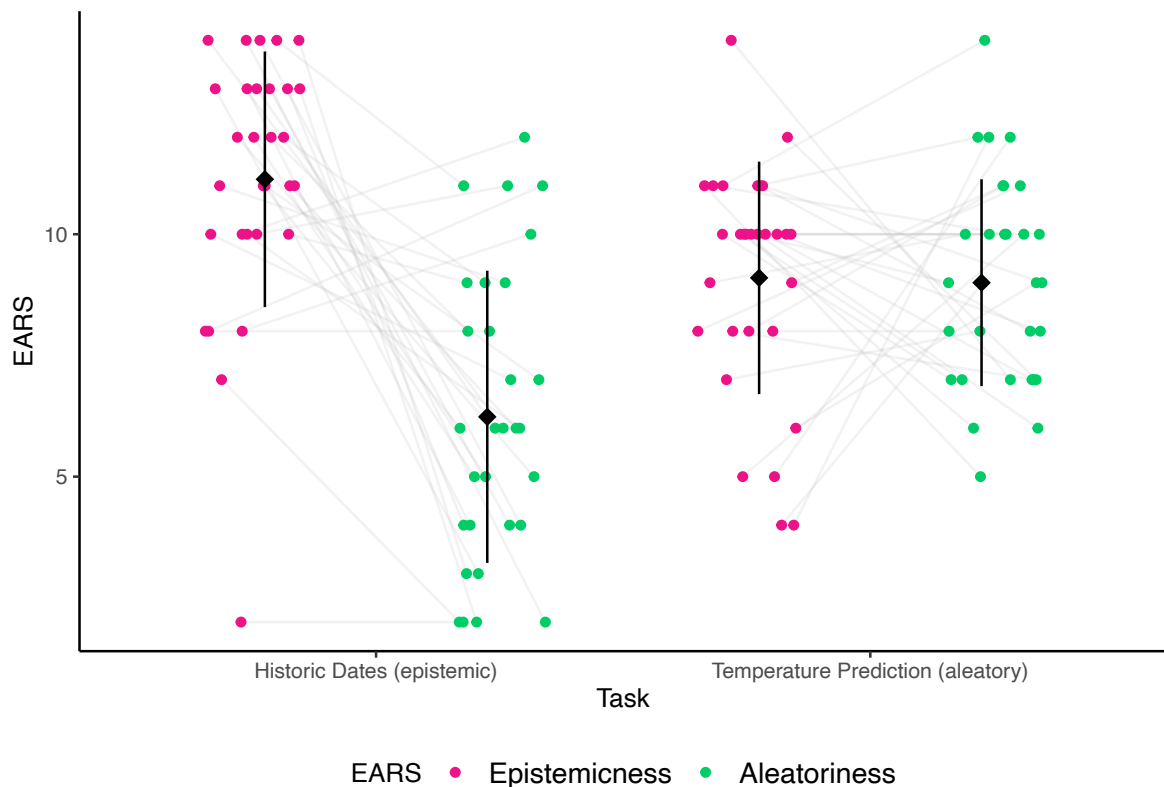
**Table 2.10** *Descriptive statistics of measures in Experiment 5*

	Historic Dates (epistemic) task		Temperature Prediction (aleatory) task	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Task Epistemicness	9.5	0.5	9.1	0.4
Task Aleatoriness	4.2	0.6	9	0.4
Epistemic Adviser Rating	3.5	0.2	3.3	0.2
Aleatory Adviser Rating	3.5	0.2	3.1	0.2
Epistemic Adviser Influence	33.3	4.4	19.0	3.7
Aleatory Adviser Influence	35.3	3.5	20.2	3.7
Adviser Distinction Score	3.5 (1.7)		4.0 (1.9)	

**Epistemic-Aleatory Rating Scale (EARS).** EARS was collected to check if perceived level of epistemic and aleatory uncertainty differed by task type. A 2 (task type) x 2 (epistemicness vs. aleatoriness) ANOVA was carried out to analyse the rating. The main effect of task type was not significant,  $F(1, 58) = 23.56, p = .382, \eta^2 = .06$ . On the other hand, the main effect of rating type was significant,  $F(1, 58) = 23.55, p < .001, \eta^2 = .20$ . This was driven by the temperature prediction (aleatory) task which was rated as having

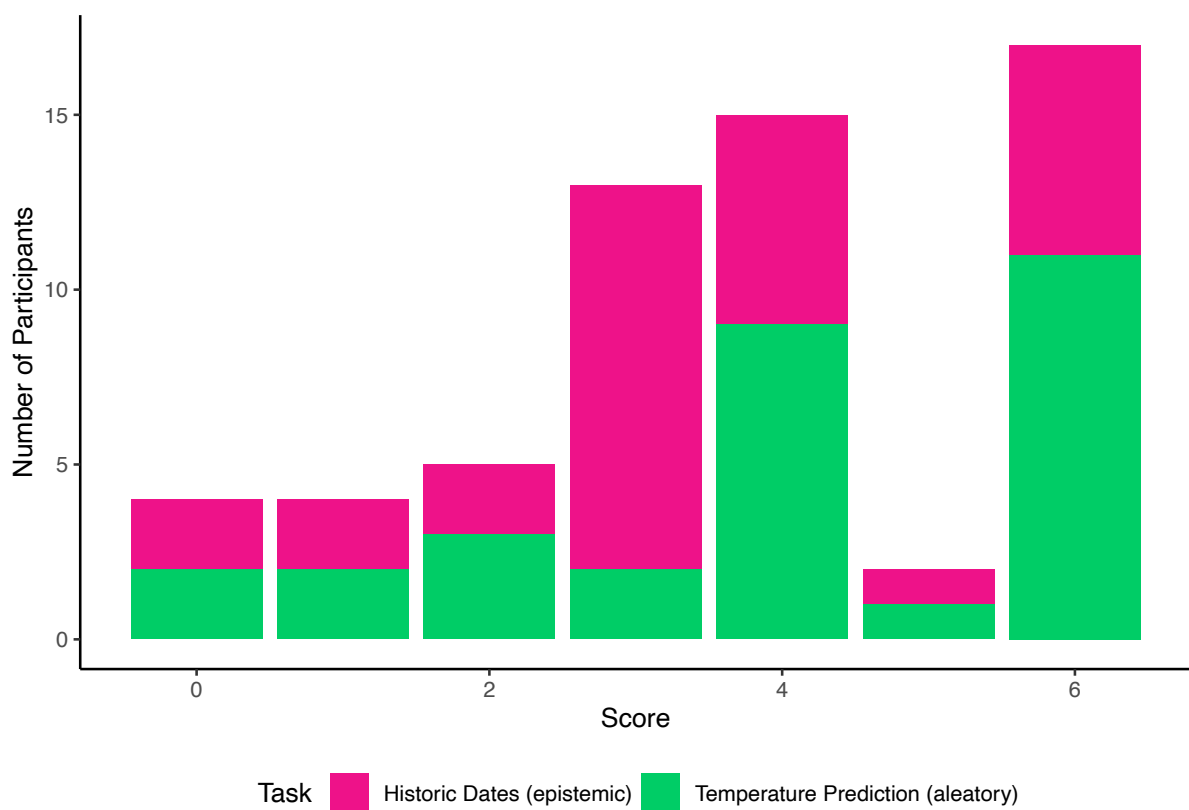
both high levels of epistemic and aleatory uncertainty The interaction effect between task type and rating type was the main focus of the analysis, which was significant,  $F(1, 58) = 21.71, p < .001, \eta^2 = .19$ . Two pairwise comparisons with Bonferroni corrections were carried out to further examine the interaction effect. The perceived level of epistemic uncertainty was significantly higher in the historic dates (epistemic) task than in the temperature prediction (aleatory) task,  $p = .003, d = -0.81$ . In contrast, the perceived aleatoriness was significantly higher in the temperature prediction (aleatory) task than in the historic dates (epistemic) task,  $p < .001, d = 1.06$ . Within the historic dates (epistemic) task, the epistemicness rating was significantly greater than the aleatoriness rating,  $p < .001, d = -1.73$ . However, the epistemicness and aleatoriness rating did not significantly differ within the temperature prediction (aleatory) task,  $p = .865, d = -0.04$ , and the numerical difference was also minimal (Figure 2.25).

**Figure 2.25** *Epistemic-Aleatory Rating Scale (EARS) by task in Experiment 5*



**Adviser distinction.** Participants were asked to identify the linguistic expressions used by each adviser to assess if participants could distinguish between the two advisers. The adviser distinction score was significantly different from three ( $M = 3.75$ ,  $SD = 1.84$ ),  $t(59) = 3.16$ ,  $p = .002$ ,  $d = 0.41$ . Many participants scored above four and only a small number of participants scoring below three (Figure 2.26). This indicates that participants were able to distinguish between the two advisers above chance level.

**Figure 2.26** Adviser distinction score by task in Experiment 5



**Accuracy.** Participants' accuracy pre-advice was calculated following the same method as Experiment 1 and 2. Specifically, the correct year of historic events and estimated future temperature was retrieved from online and compared to the participants answer. Participants' accuracy for the two tasks did not significantly differ,  $t < 1$ ,  $d = -0.15$ . A 2 (initial vs. final) x 2 (historic dates vs. temperature prediction) ANOVA found a significant main effect of time point,  $F(1, 58) = 42.9$ ,  $p < .001$ ,  $\eta^2 = .14$ . However, the

main effect of task type was not significant,  $F < 1$ ,  $p = .512$ ,  $\eta^2 = .006$ . The interaction term was significant,  $F(1, 58) = 7.04$ ,  $p = .01$ ,  $\eta^2 = .03$ . Therefore, a pairwise comparison with Bonferroni correction was carried out, which found a significant difference in final accuracy between the historic dates (epistemic) task and coin toss (aleatory) task ( $p = .040$ ,  $d = -0.54$ ). However, critically, there was no significant difference in initial accuracy between the two tasks ( $p = .062$ ,  $d = 0.15$ ), once again confirming task difficulty was similar between the two tasks.

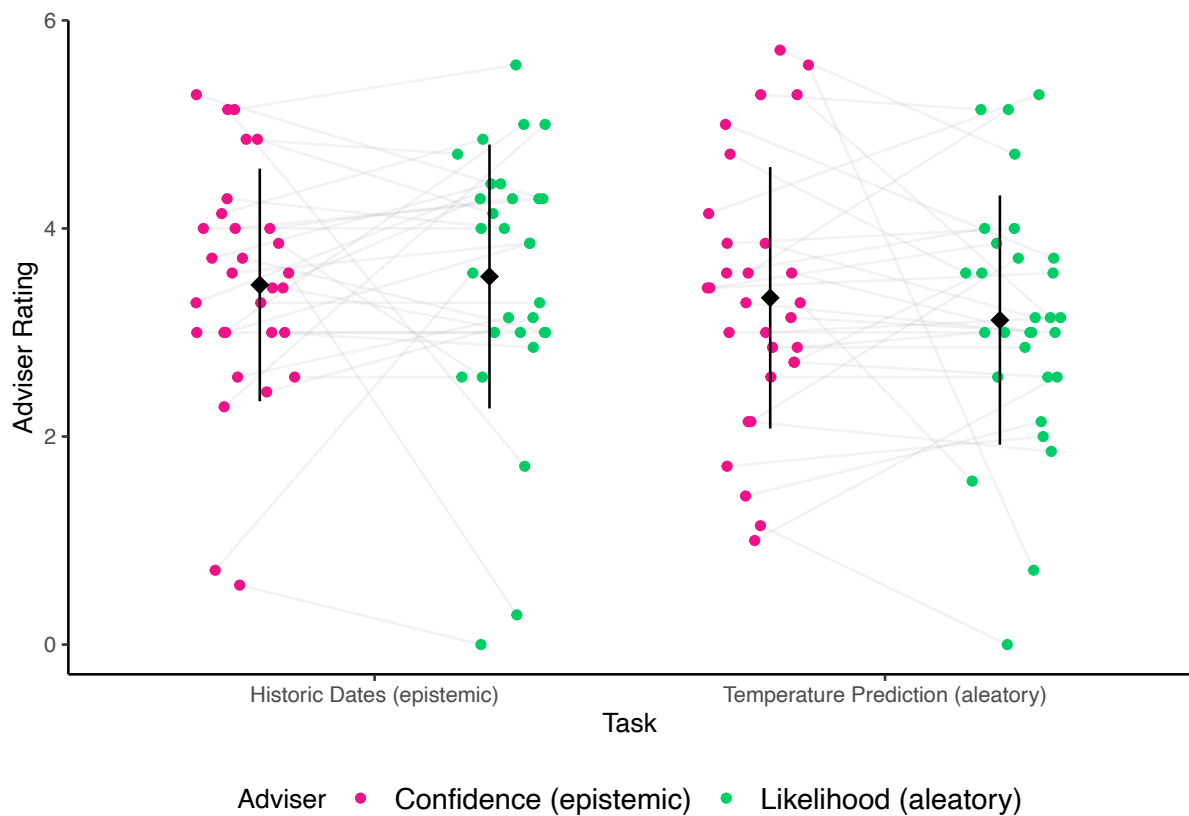
**Table 2.11** Participant’s Accuracy by Task in Experiment 5

	Historic Dates (epistemic) task		Temperature Prediction (aleatory) task	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Initial Accuracy (%)	59	0.18	62	0.15
Final Accuracy (%)	75	0.12	69	0.13

**Adviser Rating.** It was hypothesised that the adviser using confidence term (epistemic adviser) would be perceived as more competent in the historic dates (epistemic) task than in the temperature prediction (aleatory) task. In contrast, the likelihood term-using (aleatory) adviser would be rated as more competent in the temperature prediction (aleatory) task (H1 & H2). Therefore, a 2 (task type) x 2 (adviser) ANOVA was conducted with adviser rating as the dependent variable. The main effect of task type was not significant,  $F(1, 58) = 1.06$ ,  $p = .308$ ,  $\eta^2 = .01$ . However, the overall adviser ratings were numerically higher in the historic dates (epistemic) task (Table 2.8 & Figure 2.27). This is likely to be because the perceived utility of advice was lower in the temperature prediction (aleatory) task due to the inherent randomness. The main effect of adviser was also not significant,  $F(1, 58) = 0.16$ ,  $p = .692$ ,  $\eta^2 < .001$ . The main test of the hypotheses were the interaction effects, which was also not significant,  $F(1, 58) = 0.78$ ,  $p = .382$ ,  $\eta^2 = .004$ . Therefore, the hypotheses were not supported. As visible in Table 2.8 and Figure 2.26, even numerically there were little differences in the epistemic and aleatory adviser

rating within both tasks and the numerical trends were not as expected. The lack of significant results was unexpected. Therefore, a Bayesian ANOVA was conducted to examine the support for the hypotheses. There was anecdotal evidence against the main effect of task type ( $BF_{10} = 0.43$ ). On the other hand, the evidence against the main effect of adviser type was substantial ( $BF_{10} = 0.21$ ). Comparison of the model with the two main effects and full model (main effects and interaction) found anecdotal evidence against the presence of an interaction effect ( $BF_{10} = 0.37$ ). Therefore, no evidence was found to suggest that the interaction term explains the variations in adviser rating.

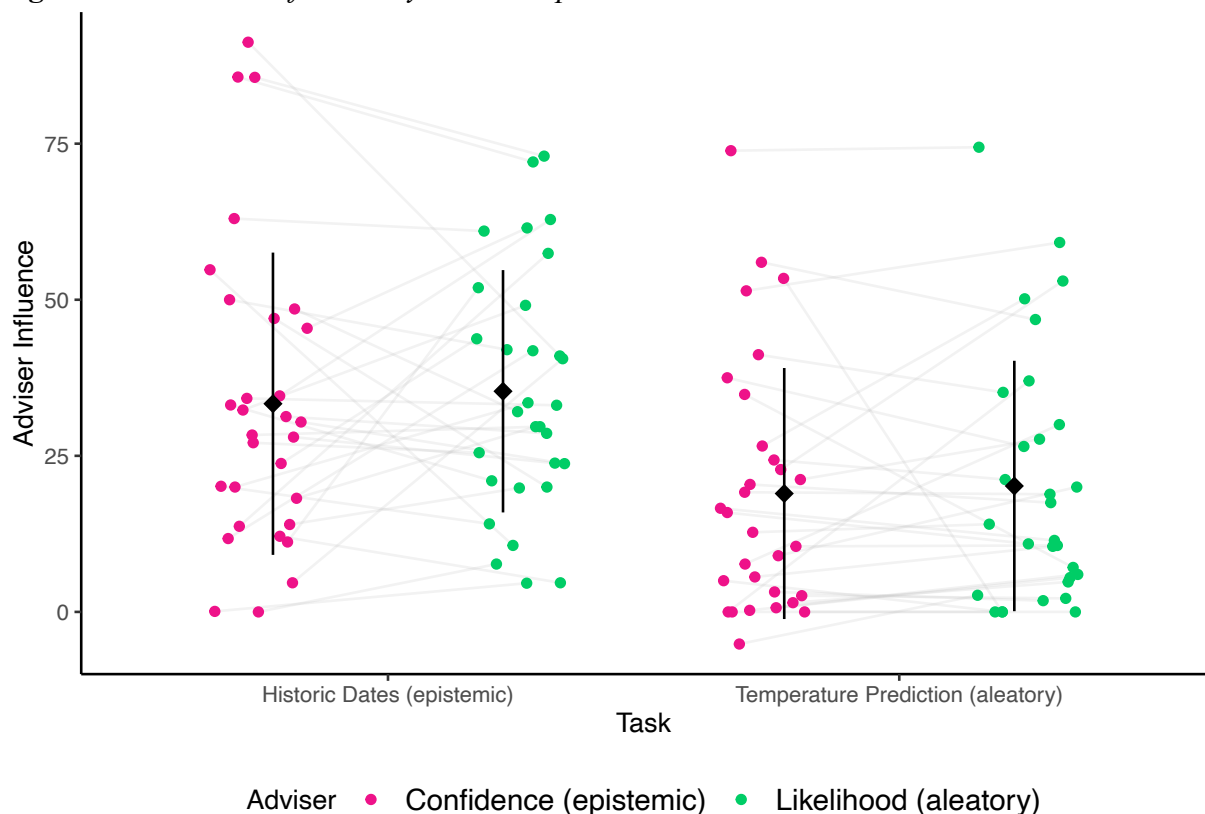
**Figure 2.27** Adviser competency rating by task in Experiment 5



**Adviser Influence.** It was also predicted that the two advisers' (confidence vs. likelihood term) influence would vary by task type (H1 & H2). To test the hypotheses, a 2 (task type) x 2 (adviser) ANOVA was carried out. The main effect of task type was significant,  $F(1, 58) = 9.39, p = .003, \eta^2 = .11$ . This indicates that participants were more reliant on the

advisers in the historic dates (epistemic) task compared to the temperature prediction (aleatory) task. Similar to the adviser rating, this is likely due to the higher aleatory uncertainty in the temperature prediction (aleatory) task. The main effect of adviser was not significant,  $F(1, 58) = 0.41, p = .523, \eta^2 = .001$ . Of critical interest was the interaction effect, which was not significant,  $F(1, 58) = 0.03, p = .875, \eta^2 < .001$ . Therefore, the hypotheses were not supported. A Bayesian ANOVA also found strong evidence in support of the main effect of task type ( $BF_{10} = 11.79$ ), confirming that the advisers were more influential in the historic dates (epistemic) task. However, there was substantial evidence against the main effect of adviser type ( $BF_{10} = 0.23$ ). The main interest was the interaction term, where there was substantial evidence for the absence of an interaction effect ( $BF_{10} = 0.27$ ).

**Figure 2.28** Adviser influence by task in Experiment 5



### 2.6.3 Discussion

In contrast to the hypotheses and the results of Experiment 2, the current experiment found minimal difference in adviser rating and adviser influence by task type. The lack of findings may be attributed to the task manipulation. For instance, participants perceived the temperature prediction (aleatory) task to have high levels of both epistemic and aleatory uncertainty. Additionally, the adviser distinction score was above chance level but still low ( $< 4$ ). The limited task uncertainty manipulation coupled with difficulties identifying the linguistic expressions used by each adviser may have led to the current findings.

However, although these factors could have contributed to the lack of findings, they are unlikely to be the key drivers of the current results. These limitations in experimental manipulations were also present in Experiment 2 but differences in perceived adviser competence and advice adoption by task were still observed. The insignificant results in the current findings suggest that there is limited impact of advisers' use of varying linguistic expressions based on the type of contextual uncertainty on advisees.

Although this was not the main analysis of interest in the current experiment, the results showed that advice adoption was significantly lower while perceived adviser competence was numerically lower in the temperature prediction (aleatory) task compared to the historic dates (epistemic) task. This pattern was also observed in Experiment 1-2 and evident in Experiment 4. This is likely because advisees perceive advice as less valuable when there are high levels of aleatory uncertainty in the advice-taking context (Walters et al., 2023). This is reasonable as highly random outcomes would not be affected by

whether or not advice was solicited. Thus, the randomness in the temperature prediction (aleatory) task might have led participants to rely less on the advisers.

## **2.7 General Discussion**

Past research has suggested that people's attitude towards uncertain advice is context dependent, where uncertain advice is preferred depending on the level of uncertainty in the advice-taking context (Du et al., 2011). The current chapter aimed to extend this notion to examine whether advisees are also sensitive to the type of uncertainty prevalent in the advice-taking context. In order to test this, epistemic and aleatory uncertainty dimensions defined by Ülkümen and colleagues (2016) were adopted. The research found that people prefer to use confidence terms (e.g., certain, sure, confident) to describe epistemic uncertainty whereas likelihood terms (e.g., probability, chance, likelihood) are favoured to communicate aleatory uncertainty. It was predicted that advisees would also prefer advisers to use context-appropriate linguistic expressions. Therefore, it was hypothesised that an adviser using confidence terms (e.g., certain, sure, confident) would be considered more competent and more influential than an adviser using likelihood terms (e.g., probability, chance, likelihood) when there are high levels of epistemic uncertainty in the advice-taking context. On the other hand, high levels of aleatory uncertainty in the advice-taking context would result in a preference for an adviser using likelihood terms over an adviser using confidence terms.

In line with Ülkümen and colleagues (2016) results, participants in the current experiments consistently distinguished between epistemic and aleatory uncertainty. The tasks where there were clear knowable answers (e.g., historic dates, total area) were rated as having high levels of epistemic uncertainty. On the other hand, tasks with inherent

randomness in the outcome (e.g., stock price prediction, coin toss) were perceived to have high levels of aleatory uncertainty. Furthermore, as the researchers suggested, participants perceive the two types of uncertainties as two separate dimensions. This was evident in the temperature prediction task (Experiments 1, 2, and 5), which was rated as having nearly equal levels of epistemic and aleatory uncertainties. In addition, people also seem to have pre-held conceptions about the uncertainties of various situations, which may be difficult to overturn. For example, in Experiment 3, there were attempts to make the stock market appear more epistemic for one group of participants (e.g., outline internal factors that affect stock price, present absolute stock price trend). However, although this group of participants perceived the stock market as more epistemic compared to their counterparts, their perceived aleatory uncertainty was still numerically higher than the epistemic uncertainty rating. This suggested that participants already perceived the stock market as aleatory and the task manipulations were not sufficient to change their beliefs.

The main hypotheses of the experiment predicted that advisees would perceive advisers who use context-appropriate linguistic expressions as more competent and adopt their advice more than advisers who do not use such linguistic expressions. Overall, only Experiment 2 yielded significant results, but the other experiments showed limited support for the hypotheses. More specifically, Experiment 2 found that likelihood term-using (aleatory) adviser was perceived as more competent in the temperature prediction (aleatory) task than in the historic dates (epistemic) task. The difference in confidence term-using (epistemic) adviser did not significantly differ by task but the numerical trend was as hypothesised. Additionally, confidence term-using (epistemic) adviser was significantly more influential in the historic dates (epistemic) task than in the temperature prediction (aleatory) task. In contrast, although the difference was not significant, the

likelihood term-using (aleatory) adviser was numerically more influential in the temperature prediction (aleatory) task. Furthermore, the numerical trend of adviser rating and adviser influence within the tasks were in the expected direction. In addition, there were numerical differences in adviser influence within the aleatory task in Experiment 3 and 4. Moreover, replication of Experiment 2 (i.e., Experiment 5) did not yield the same results. The findings overall suggest that people might prefer advisers to use context-appropriate linguistic expressions, but this is a weak preference that is unlikely to make a meaningful difference in adviser evaluation and advice adoption in real-life. This is also supported by the adviser distinction score, where participants often did not perform above chance level and scores were low in all experiments (i.e., below 4). This indicates that participants were unable to recognise that the two advisers were using different linguistic expressions. Participants would have formed stronger impressions of each adviser if they were concerned about the advisers' linguistic expressions. In sum, the experimental results suggest that people are largely insensitive to the relationship between the type of uncertainty in the advice-taking context and adviser's linguistic expressions used in their advice.

On the other hand, the experimental results should not be taken as evidence that advisees are impervious to the level of epistemic and aleatory uncertainty in the advice-taking context. In particular, increase in perceived level of aleatory uncertainty decreased participants reliance on the advisers. For example, in Experiments 4 and 5, the overall adviser rating and influence was lower in the task designed to have high levels of aleatory uncertainty. Similar patterns were observed in previous research where people were less likely to seek financial advice when they perceive the stock market as high in aleatory uncertainty (Walters et al., 2023). This is expected as increased randomness would reduce

the usefulness of advice. Therefore, the current findings hint that contextual factors may influence general advice-taking behaviour. Building on this, future research could investigate if the advice-taking context influences advisees' motivation for advice-taking. For example, Schrah and colleagues (2006) found that advisees may request advice to test if their answer is correct or to aid decision-making. The former type of advice-taking may be more prevalent under high aleatory uncertainty, where advisees seek for agreeing advice to boost their confidence.

Although this was not directly tested in the current experiments, the results contribute to the literature on uncertain advice. In the experiments, one of the advisers used words such as probability (i.e., aleatory adviser), which signals uncertainty (Du et al., 2024; Yates et al., 1996). However, the likelihood term-using (aleatory) adviser was perceived more competent and was more influential in some experiments (e.g., Experiment 2, 3). Previous findings showed that advisees prefer certain advisers and tend to disregard uncertain advice (Price & Stone, 2004; Schultze & Loschelder, 2020). However, few studies have found that people often prefer and adopt less certain advice. For example, Gaertig and Simmons (2018) found that people are equally likely to adhere to confident advice and advice that incorporates some level of uncertainty. In a more recent study, advisees preferred advisers gave their answer in small ranges compared to advisers who conveyed higher certainty through exact point estimates. Albeit extremely uncertain advice backfires and reduces advice adoption (Du et al., 2024). Additionally, investors are more likely to invest in firms that provide their predicted performance in range estimates compared to firms that give exact numbers (Du et al., 2011). The willingness to accept uncertain advice may be driven by the evaluation of the advice. Advisees tend to perceive advice given as a range as more accurate, informative and credible (Du et al., 2011). On

the other hand, point estimates appear overly precise which in turn harms the credibility of the adviser (Loschelder et al., 2016; Schultze & Loschelder, 2020). This may be because an adviser's high confidence could be construed as overconfidence. Advisees have been found to mistrust overly confident advisers and even penalise them (Stanciu & Fiser, 2022; Vullioud et al., 2017). On the other hand, researchers have also suggested that people prefer confident advisers as they are assumed to have more information and are more competent (i.e., confidence heuristic) (Price & Stone, 2004; Thomas & McFadyen, 1995). Additionally, uncertain advice is often perceived to be of lower quality (Schultze & Loschelder, 2020).

The mixed findings on uncertain advice in the literature could be explained by examining the tasks used in previous studies. Research in judgement and decision-making has often used tasks where there is a correct answer (Bonaccio & Dalal, 2006), which is a characteristic of epistemic uncertainty. In contrast, tasks that have found preference for uncertain advice generally used tasks where the outcome is more random such as, sports game or predicting the stock market (e.g., Gaertig & Simmons, 2018). The high levels of aleatory uncertainty in these advice-taking contexts may have resulted in preference for uncertain advice. Indeed, Budescu and Wallsten (1995) have argued that people are sensitive to the congruency between the type of uncertainty in events and the description of the events. This is likely to extend to advice-taking, where advisees evaluate the uncertainty in the advice-taking context and prefer advice that matches their evaluation. However, this has not been directly tested in the current experiments and conclusive claims cannot be made. Nonetheless, there is a possibility that context accounts for some of the variations in previous findings on uncertain advice and would be an interesting area for future research to explore.

### **2.7.1 Limitations**

The current experiments had several methodological limitations. Firstly, there were challenges with designing tasks that were distinctly high in epistemic or aleatory uncertainty. For example, it was difficult to design a task that was solely high in aleatory uncertainty as people are less likely to seek advice when there are high levels of aleatory uncertainty in the advice-taking context (Walters et al., 2023). As a result, the perceived level of epistemic and aleatory uncertainty often did not differ within the tasks that were designed to be high in aleatory uncertainty (e.g., temperature prediction task in Experiments 1, 2, and 5). Experiment 4 used a coin toss task, which was rated as having high levels of aleatory uncertainty and low levels of epistemic uncertainty. This led participants to disregard advice which made it infeasible to compare across tasks and to test the hypotheses. However, it is worth noting that the task manipulation is unlikely to be the main driver of the lack of findings in the current chapter, considering that insignificant results were observed despite successful task manipulation in some experiments.

Second, it is unclear if participants were unaffected by the advisers varying linguistic expressions or simply were unaware that the two advisers were using different expressions. In the current experiments, advice was given in text form and the different linguistic expressions were given before the advisers' answer to the question. Therefore, participants might have focused on the advisers' answer rather than their linguistic expressions. The adviser distinction score was no better than chance level in two experiments and the average score was below four in the remaining three experiments. It is possible that the preference for different linguistic expressions in advice arise upon deliberate reflection. In Ülkümen and colleagues' (2016) research, participants carefully

evaluated between the two expressions for a given scenario. Therefore, the results of the current experiments may differ if advisees were required to reflect on the two different expressions and better distinguished between the two advisers. In the current experiments, some efforts were made to aid recognition of the two linguistic expressions, such as capitalising the expressions and making advice appear line by line. However, other methods such as ensuring participants have learned the different expressions prior to the test trials could yield different results. This method was not employed in the current experiments as advisees would rarely evaluate advice with such efforts in real-life. Nonetheless, preference in advisers to use context appropriate linguistic expressions may arise when advisees are asked to deliberately choose between two expressions.

Another limitation may be that the type of uncertainty prevalent in the tasks were apparent, and thus participants did not feel the need to examine the advisers' linguistic expressions to determine the contextual uncertainty in the task. In most of the tasks, it was clear whether there is high epistemic or aleatory uncertainty. For example, in the historic dates task it was evident that greater knowledge in the area or gathering further information would resolve the uncertainty. Therefore, advisees may have focused on the advisers' answers rather than trying to gather further information about the advice-taking context from advisers' linguistic expressions. Designing a task where the epistemic and aleatory nature of the task is ambiguous could prompt participants to attend to the advisers' linguistic expressions. This was not the focus of the current experiments, as the aim was to investigate whether advisers' linguistic expressions influence advice-taking in naturalistic contexts. Future research could manipulate the ambiguity of the advice-taking context (e.g., rather there is high epistemic or high aleatory uncertainty is unclear) could clarify the association between different types of uncertainty and linguistic expressions.

### **2.7.2 Conclusion**

In five experiments, the relationship between contextual uncertainty and advisers' linguistic expression of confidence was investigated. It was hypothesised that when there are high levels of epistemic uncertainty in the advice-taking context, advisees would prefer advisers to use confidence (certain, sure, confident) terms over likelihood terms (probability, chance, likelihood) terms in their advice. In contrast, when there are high levels of aleatory uncertainty, an adviser who uses likelihood terms would be evaluated more positively and would be more influential. Only one experiment supported the hypotheses and minimal difference in the two advisers were found in the remaining experiments. Therefore, advisers' varying linguistic expressions are likely to have minimal effect on adviser evaluation and advice adoption. However, there were methodological limitations, and different results may be obtained in future research that uses different tasks or require deliberate reflection of advice. Additionally, the experiments found hints that the type of uncertainty in the advice-taking context impacting general advice-taking behaviour. For example, perceived adviser competence and advice adoption decreased with increased aleatory uncertainty in the advice-taking context. Furthermore, the experiments also contribute to previous literature on uncertain advice. The use of likelihood terms in advice has previously been perceived as an expression of uncertainty and often reduced advice adoption. However, in the current experiments, participants were equally likely to adopt advice from the confidence term-using adviser and the likelihood term-using adviser, and sometimes higher for the likelihood term-using adviser. The type of contextual uncertainty may provide clues for understanding the ambiguous findings in the uncertain advice literature.

# Chapter 3: External Information Ambiguity and Adviser Confidence

---

## 3.1 Introduction

Life is filled with moments that require decision making, from small daily decisions such as what to eat to life-altering decisions such as which career path to take. According to Bayesian decision making theory, people use their existing knowledge to evaluate evidence and incorporate this evaluation into their knowledge (Gold & Shaplen, 2007; Kyburg 1987; Nassar et al., 2010). In addition to evaluating evidence that is observed firsthand, people also seek out social information in the form of advice. Advice can come from many sources, from close others to professionals such as doctors, lawyers or financial advisers. Advice-seeking occurs due to various reasons, such as a lack of relevant knowledge or experience, or difficulties in evaluating the evidence. In Bayesian decision making, advice is considered as simply another piece of information (Pescetelli et al., 2021). However, some researchers claim that advice should not be treated as another piece of information due to its unique impact on our decision-making (Milyavsky & Gvili, 2024). Advice carries a social aspect, where there are costs of ignoring advice (Rader et al., 2017). When their advice is rejected, advisers negatively judge the advisee's competency and are unwilling to provide future advice (Belkin & Kong, 2018; Blunden et al., 2019; Palmeira & Romero Lopez, 2023). This would influence how much advice is used compared to other non-social information (Sah et al., 2006; Rader et al., 2017). In light of the frequency people seek advice, and the particular effect advice has on decision-making, there has been considerable research interest in characterising how people seek and use advice during decision-making (Bonaccio & Dalal, 2006; Kämmer et al., 2023).

There are countless advisers in the world, and thus it is essential to determine the reliability of the advice on offer. Some indicators of reliability are readily available such as certificates or professional qualifications (e.g., doctors, lawyers). However, sometimes the reliability of the advice source may not be accessible or straightforward. Therefore, people often use how confident an adviser is to determine the quality of advice. There are various definitions of confidence but in this chapter, confidence will be defined as the expectation or expression of the extent to which a decision, opinion, or recommendation is correct (Klayman et al., 1999). Researchers repeatedly find that people prefer more confident advisers, and expressing higher confidence promotes greater advice adoption (Bonaccio & Dalal, 2006; Pulford et al., 2018; Sniezek & Van Swol, 2001).

The question remains, what makes confident advisers so influential? This is because adviser confidence is processed in relation to diverse factors. For instance, people use a *confidence heuristic* where advisees assume that confident advisers are more knowledgeable or hold more reliable information (Price & Stone, 2004; Thomas & McFadyen, 1995). People also deem confidence as a sign of how committed the adviser is to their advice (Vullioud et al., 2017). In addition, people are sensitive to the relationship between confidence and accuracy. An adviser's inability to judge their confidence accurately based on their expected performance (i.e., overconfidence) leads to punishment or mistrust in the adviser (Sah et al., 2013; Tenney et al., 2008; Tenney et al., 2019; Van Zant, 2022) even in domains unrelated to the advice-taking context (Vullioud et al., 2017).

Despite the various inferences people make using adviser confidence, how people process adviser confidence in relation to the external information has received limited attention.

In fact, the difference in advice-taking context has largely been ignored as it was assumed to have minimal impact on how people process adviser confidence (Bonnacio & Dalal, 2006). Yet, contextual factors influence people's approach to decision-making. For example, Billings and Scherer (1988) suggest that people's decision-making strategy differs depending on whether or not they are required to give an answer. When people are simply asked to evaluate available evidence, people search for more information compared to when they are asked to make a decision. This difference in strategy derived from difference in decision-making context also suggests that how adviser confidence is processed may be context dependent. In support of this notion, Gaertig and Simmons (2018) found that people do not always prefer confident advisers. Instead, advisers who express uncertainties using probabilistic terms or by giving range of outcomes are sometimes favoured. The tasks used (e.g., sports game outcome, future stock price prediction) in these studies had inherent uncertainty where the outcome could be affected by multiple factors. Therefore, advisers who incorporated this contextual uncertainty in their advice were more trusted and preferred. Similarly, online reviews that express doubt are more trusted compared to reviews that are unequivocally certain (Evans et al., 2011). These studies taken together demonstrate that advisees take into account subtle features of advice and how advisers express uncertainty and confidence. The experiments reported in this chapter extend this idea to explore how advisees process the discrepancy between adviser confidence and external information ambiguity. The current chapter aims to increase the understanding of how people assess external information during decision-making and the role adviser confidence plays in information evaluation.

One impact of unexpected adviser confidence would be its influence on advisees' confidence. Consider the first case where a decision maker has evaluated external

information that seems ambiguous (i.e., evidence not clearly favouring one choice over other options). In this case, people would not undoubtedly favour one choice over the other and the perceived difficulty of the decision would increase as it would be unclear how to combine and assess the evidence. This difficulty in choice may result in greater dependence on advice (Soll & Larrick, 2009). Under this condition, advisees would be surprised if the adviser is highly confident. How advisees process this unexpected confidence may depend on adviser expertise. For instance, advisees might assume that the expert's high confidence, despite ambiguous evidence, stems from greater knowledge or access to information (i.e., confidence heuristic). Therefore, when an expert adviser agrees with their opinion, people would feel more confident as it is evidence supporting their opinion (Julian et al., 1968; McGarty et al., 1993; Schultze et al., 2015). On the other hand, advisees would decrease their confidence when an expert disagrees (Julian et al., 1968; McGarty et al., 1993; Schultze et al., 2015).

In contrast, an adviser who appears similar to themselves (i.e., novice) but communicates high confidence would appear overconfident. Such advisers (i.e., whose confidence is poorly calibrated to their objective accuracy) are mistrusted and even punished (Rader et al., 2017; Sah et al., 2013). Although, the advisers' accuracy information is not directly available in the current experiments, people are able to infer task difficulties using their metacognition (Efklides 2006; Efklides 2014). Therefore, based on the assessed task difficulties and adviser expertise, people might also deduce an adviser's expected accuracy. Hence, when a novice adviser is overly confident, advisees might assume that the novice lacks the ability to assess their own accuracy and are less affected by their advice. Therefore, change in confidence following advice would be weaker when

receiving advice from a novice (Soll et al., 2021). Thus, the following hypothesis is laid out:

H1. When the external evidence is ambiguous, but the adviser displays high confidence, people will follow the adviser and increase their confidence when an adviser agrees with their initial answer. On the other hand, when an adviser disagrees with their initial answer, advisees will decrease their confidence. This effect will be greater for an expert adviser compared to a novice adviser.

Hypothesis 1 is an intuitive and relatively straightforward prediction. A less obvious case holds when a decision maker has reviewed evidence that seems to point clearly and unambiguously to a particular decision, but the adviser expresses high degrees of uncertainty. Consider a scenario where an expert agrees with the advisee's initial decision, but with much less confidence than the advisee themselves. According to previous research, advisees should increase their confidence in their initial decision based on agreement from the adviser. A simple Bayesian model of advice integration would predict the same outcome. An advisee's confidence would only increase after observing evidence that supports their decision (i.e., agreeing adviser) as it leads to higher probability of being correct. On the other hand, unexpected information grabs our attention hence it is processed more in depth (Horstmann, 2015; Soltani & Izquierdo, 2019). Advisees may be surprised by the adviser's low confidence and conclude that they had been overconfident. This pattern would be observed regardless of adviser expertise, but the effect would be greater with an expert adviser considering the positive correlation between expertise and adviser influence (Birnbaum & Stegner, 1979; Harvey & Fischer, 1997). Therefore, the following hypothesis will be tested:

H2. When external evidence is unambiguous, but an adviser is not confident, people will follow the adviser and decrease their confidence even when the adviser agrees with them. This effect will be greater for an expert adviser compared to a novice adviser.

Hypotheses 1 and 2 build on the notion that unexpected adviser confidence raises doubt about the advisee's own confidence. When there is a large difference between an advisee's and an adviser's confidence, it would be natural for advisees to question their judgement and their evaluation of the evidence. Taking a step further, this could prompt advisees to analyse why they were overconfident and conclude that their decision-making strategy was incomplete or inaccurate. If advisees conclude that they are misevaluating the evidence, they might try to deduce and adopt the adviser's decision-making strategy. Adviser confidence could be a useful tool for this, by considering evidence strength in relation to adviser confidence. For example, if one predicts an outcome with high confidence based on a particular piece of evidence, but the adviser displays low confidence, one could infer that this evidence source is not strongly diagnostic. Over repeated interaction, advisees would be able to deduce how predictive each piece of information is and change their decision-making strategy accordingly. Indeed, studies on cue learning has found that people are able to infer the relationship between cues from accuracy feedback (Droll et al., 2009; Little & Lewandowsky, 2009). Explicit feedback was not given in the current experiments, but changes in evaluation and use of evidence may still occur. Therefore, in addition to the pre-registered hypotheses, the following exploratory hypothesis is proposed.

H3. The weight people assign to each piece of evidence will change over time following repeated interaction with the advisers.

### 3.1.1 Research Outline

Three pre-registered experiments (Experiments 6-8) were carried out to test the above hypotheses. In all experiments, the task was to predict basketball game outcomes. The games used in the experiments were taken from United States (U.S.) National Basketball Association (NBA) games, but the team names were masked. Four basketball statistics with varying predictive power were provided to aid prediction. There were also two advisers, one expert and one novice. The expert's weighting of the four basketball statistics was well aligned with their objective reliability whereas the novice would over- or under-weight some statistics. Estimation tasks in the form of sports game outcome prediction have been widely used in studies of advice-taking (Bonaccio & Dalal, 2006). This naturalistic task resembles real-world advice-taking situations where the outcomes are unclear and determined by multiple factors with varying degrees of predictive validity regarding the outcome.

In the task, participants made their game outcome predictions on a continuous scale that simultaneously indicated their predicted winning team and their confidence in this prediction. Along other benefits, continuous measures allow greater precision in capturing advice adoption (Sah et al., 2013). In the experiments, the focus was to see how much people are willing to follow advice in relation to their own decision and external information evaluation. Therefore, communication was one sided and advisees were aware that their degree of advice adoption would not be communicated to the adviser. This set-up was to prevent social expectations surrounding advice adoption from resulting in greater adviser influence (Kämmer et al., 2023; Rader et al., 2017). Advice was given in text form for simplicity and for ease of processing (Van Swol & Sniezek, 2005).

All experiments followed the standard Judge Adviser System (JAS) (Sniezek & Buckley, 1995), where participants first reported their initial answer and then were provided with advice. After, participants were given the option to update their answer. In Experiment 6, participants and advisers gave their answer using the same format. Specifically, participants and advisers gave their predicted winning team and accompanying confidence as a specific numeric value (e.g., 65%). This method allows for precise measure of advice adoption. However, participants may also simply change their confidence to match the advisers' confidence (Friedemann et al., 2024) rather than reporting their true confidence. Therefore, in Experiment 7 advisers gave their confidence using five linguistic expressions. In Experiment 8, the adviser gave their confidence through numbers, but participants were asked to indicate their confidence through five linguistic expressions. Linguistic expressions are generally associated with a range of confidence values (e.g., 50-59%). Therefore, there is no specific confidence value for participants to match their confidence. Hence, it is likely that change in confidence in these experiments reflect participants' true confidence rather than a result of confidence matching.

### **3.2 Experiment 6**

Experiment 6 examined how advisees process the discrepancy between adviser confidence and external information ambiguity. More specifically, cases where external information seemed ambiguous, but an adviser was highly confident, or an adviser was unexpectedly unsure even though the evidence appeared clear were investigated. Additionally, the role of adviser expertise was also examined, where participants viewed an expert and a novice adviser. It was expected that advisees would flexibly adjust their confidence and adjust their use of the external information. This was tested through a basketball game outcome prediction task.

### 3.2.1 Methods

**Participants.** Participants were recruited through Prolific (www.prolific.com). In total, 30 participants (14 = Female, 16 = Male,  $M_{age} = 35.6$   $SD_{age} = 9.8$ ) took part in Experiment 6. Prior to starting the experiment, all participants provided informed consent in accordance with the University of Oxford Central University Research Ethics Committee guidelines. Participants were compensated £9/hour based on the median duration of the experiment.

**Task.** The task adopted the judge-adviser system (JAS) (Sniezek & Buckley, 1995). In this experiment, a binary decision-making task was used where participants were asked to predict the outcome of basketball games. The basketball game data was from real National Basketball Association (NBA) games, collected from the NBA website (<https://www.nba.com>) on February 2020. However, the names of the teams were masked, and they were simply referred as ‘Home Team’ or ‘Away Team’. Participants were presented with some basketball statistics information. Specifically, the effective field goal rate (eFG%), offensive rebound rate (ORB%), points scored the previous month, and percentage of wins this month was given. The basketball statistics were explained in layman’s term and thus, prior knowledge in basketball was not required. The explanation of the basketball statistics is shown in Figure 3.1.

#### **Figure 3.1** *Explanation of the four basketball statistics*

**Shooting accuracy (eFG%):** Percentage of successful shots by the team (adjusted for the difficulty of the shot).

**Offensive rebound (ORB%):** Percentage of time the team retrieved the ball when it bounced off the hoop after their own missed shot.

**Points scored last month:** Average points scored by the team across all of their games last month.

**Percentage of wins this month:** Number of matches won by the team this month compared to the number of matches they played.

**High numbers are desirable for all 4 measures.**

For some trials, all four statistical measures indicated that one team would win (i.e., unambiguous external information). Other times, two measures favoured one team, and the remaining two measures favoured the other team (i.e., ambiguous external

information). For example, eFG% and points scored last month would be higher for Home Team whereas offensive rebound rate (ORB%) and percentage of wins this month would be higher for the Away Team.

Participants were also provided with advice to aid decision-making. There were two advisers, one expert and one novice. Both were virtual advisers rather than based on predictions of any real individuals. Specifically, the expert adviser's predictions were generated by running a linear regression on the collected NBA game data, from a dataset comprising of 168 games. First, the difference in statistics data between Home Team and Away Team was calculated (Home Team – Away Team). Second, a linear regression was carried out with the difference in game score as the dependent variable. The independent variables were the difference between Home Team and Away Team of the four predictors (i.e., basketball statistics) defined above. These four predictors were selected based on an initial regression model which had these four plus a further three basketball statistics as predictors. From this set of 7 predictors, Effective field goal (eFG%) was selected as it had the largest regression coefficient. Two further variables with the smallest positive regression coefficients but still significant p-values were selected: ORB% and percentage of wins this month. Lastly, one non-significant variable (win percentage) was selected to reduce the regression model's predictive power. The adjusted  $R^2$  was .95 with all seven variables but reduced to .80 in the final (4 predictor) model. The prediction of the novice adviser was created by manipulating the coefficients of the four selected predictors. The novice adviser undervalued eFG% and points scored the previous month (coefficients divided by half) but overvalued ORB% and percentage of wins this month (coefficients multiplied by two) (Table 3.1). Using the regression equation, the advisers predicted score difference between the two teams (home vs. away) was generated for each game based

on the actual basketball statistics information. The adviser’s predicted score difference was then used to calculate the probability of the Away Team winning following the equation below. Here, the  $x$  values were the adviser’s predicted score difference while  $\beta$  was 0.1. The  $\beta$  value was set to have to most reasonable probability range as any value lower led to conservative prediction while higher values resulted in overconfidence.

$$P(y = 1) = \frac{1}{1 + e^{-x*\beta}}$$

These probabilities were calculated separately for the expert and the novice adviser. The probabilities were then translated to be above 50%. For example, probability of 20% that Away Team would win was re-written as 80% probability that Home Team would win.

**Table 3.1** Regression Coefficients for Expert and Novice advice

Predictors	Expert		Novice	
	Regression Coefficient	Standardised regression coefficient	Regression Coefficient	Standardised regression coefficient
eFG%	123.53	0.78	61.77	0.39
ORB%	0.32	0.23	0.65	0.46
Points Scored Previous Month	0.12	0.05	0.06	0.03
Win% This Month	6.49	0.12	12.99	0.23

Descriptions of the two advisers were also provided to manipulate participants’ perception of expertise of the two advisers. The expert adviser was described as someone who makes mostly accurate decisions based on current news and team data, whereas the novice was described as a basketball fan who sometimes guesses correctly. Two photos from the Chicago Face Database (<https://www.chicagofaces.org>) were used in this experiment to represent the advisers. The faces selected were matched regarding gender, race, attractiveness and trustworthiness to prevent these factors from influencing advice adoption. Under the photos, the advisers were labelled as either ‘Expert’ or ‘Novice’.

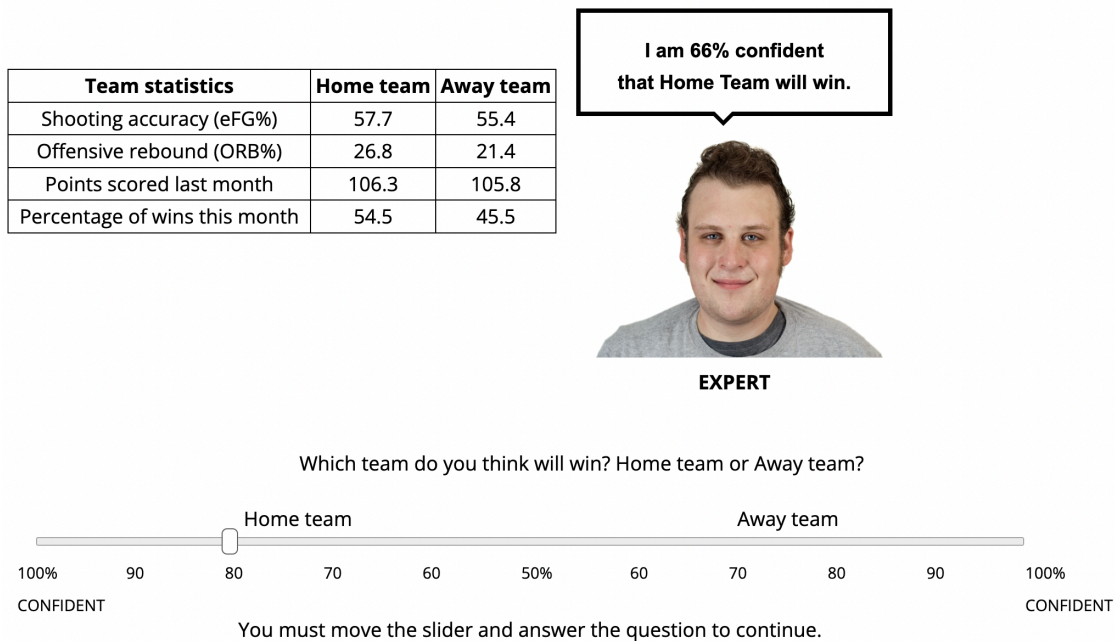
Lastly, the speech bubble for each adviser also varied, where one had round edges but the other was square. The adviser assignment (as expert or novice), presentation order, photos used, and speech bubble shape were all randomised. Participants viewed each adviser equal number of times.

The purpose of the study was to examine how participants react to and learn from unexpected adviser confidence. Therefore, out of 168 games, 24 games were selected that exhibited particular relationships of interest between adviser confidence and the ambiguity of the four basketball statistics. More specifically, on certain critical trials, the adviser displayed low confidence ( $< 72$ ,  $M = 66$ ) when the external information was consistent (i.e., all four basketball statistics favoured one of the two teams) or expressed high confidence ( $> 77$ ,  $M = 81$ ) when the external information was inconsistent (i.e., two of the statistics favoured one team and two favoured the other). In addition to these critical trials, there were 24 “filler” trials where the level of external information ambiguity matched adviser confidence (i.e., adviser displays high confidence with unambiguous evidence or low confidence with ambiguous information). Therefore, there were eight types of trials which differed in terms of information consistency (information unambiguous vs. ambiguous), adviser type (expert vs. novice) and adviser confidence (high vs. low). There were 48 experimental trials in total and among these 24 were critical test trials. In half the trials, advice was given by the expert adviser and in the remaining half a novice adviser appeared.

Participants’ answers were collected using a horizontal continuous scale ranging from 100% Home Team to 100% Away Team. The slider starting point was the mid-point (50%). Participants answered by moving the slider along the horizontal scale. Thus,

participant gave their answer and confidence level simultaneously. The midpoint, 50% confident, was un-selectable to force a decision between the two teams. Once participants submitted their initial prediction, they were given an adviser's answer and confidence level. This information remained visible while participants moved the slider to register their final prediction about which team would win. The slider starting point was once again the mid-point. An example of a trial where the four basketball statistics are all higher for Home Team, but an expert adviser displayed low confidence is shown in Figure 3.2.

**Figure 3.2** Example question of unambiguous evidence trial with expert adviser in Experiment 6



**Procedure.** Participants first received instruction regarding the task and read the descriptions of the two advisers. Next, an example question was given along with definitions of each of the four basketball statistics. Participants then answered the example question based on the four presented statistics before receiving advice from both the novice and the expert adviser simultaneously. The two advisers gave the same answer (i.e., same confidence and winning team prediction). Participants then registered a final

answer indicating their predicted winning team and their associated confidence for the example question.

Participants then proceeded to a three-trial training block. Each trial comprised of an initial stage where participants viewed team statistics and gave their prediction and confidence. After, participants viewed advice from both advisers together and gave their answer using the same scale. Once participants provided their final answer, the actual game outcome—the winning team and the score difference between the two teams—was revealed. In the first two trials, the two advisers predicted that the same team would win but their level of confidence varied. In the first trial, the expert adviser was more confident compared to the novice adviser and vice versa in the second trial. In the last trial, the expert and novice predicted opposing teams with the expert being slightly more confident than the novice. The expert was correct. This trial was designed to reinforce that the expert was more accurate.

Upon completing the training block, participants received instructions about the test block explaining that they would only receive advice from one adviser on each trial (either the expert or the novice) and that they would no longer receive game outcomes. The explanation of each basketball statistics (effective field goal rate (eFG%), offensive rebound rate (ORB%), points scored the previous month, and percentage of wins this month) was once again provided to help refresh participants' memory.

The 48 experimental trials were split into two blocks to prevent fatigue and provide a chance to rest between the two blocks. There were equal numbers of each trial type (8 trial types in total) in both test blocks. The same adviser appeared for three consecutive

trials and then alternated in 3-trial sequences. Which adviser was shown first was counterbalanced.

Once participants completed the test blocks, their rating of competence for each adviser (expert and novice) was collected. The questionnaire, adopted from Gaertig and Simmons (2018) (Appendix A), comprised of 7 questions (e.g., “How knowledgeable is this adviser?”, “How credible is this adviser?”) and participants answered on a 7-point Likert-scale (1 = *Not at all*, 7 = *Extremely*). The average score of the 7 questions was taken as the advisers’ competence rating. Participants were also asked to estimate each advisers’ use of the four basketball statistics. Participants ranked each basketball statistics from used the most to the least, separately for the expert and the novice adviser. Which adviser participants ranked first was counterbalanced. Finally, participants reported their age, gender and basketball knowledge. Basketball knowledge was collected on a 5-point Likert-scale (1 = *None*, 5 = *Expert*). All participants were debriefed after completing the experiment.

**Data Analysis.** The current experiments aimed to explore how people process external information and adviser confidence. In particular, the mismatch between adviser confidence and external information ambiguity was researched: trials in which the adviser was highly confident despite ambiguous external information (i.e., basketball statistics), and trials in which the adviser exhibited low confidence even though all basketball statistics suggested a clear winning team. There were 12 trials with ambiguous external information and high adviser confidence (referred as ambiguous evidence trials from here on) and 12 trials with unambiguous external information and low adviser confidence (referred as unambiguous evidence trials from here on), half of which involved the expert

adviser, and the other half involved the novice adviser. For each type of trial, the aim was to examine if advice influenced participants' prediction of the winning basketball team and accompanying confidence. Participants' answer was collected on a scale ranging from 100% confident that the home team would win to 100% confident that the away team would win, while the mid-point (50% confident) was unselectable. Therefore, participants' confidence ranged from 51% to 100% confident for each team. Thus, when a participant became more confident in their initial answer, the *adviser influence* measure would be positive (with a maximum value of +49). In contrast, if a participant became less confident or changed their answer after receiving advice, adviser influence would be negative (with a maximum value of -100, if the participant went from being 100% sure one team would win to 100% sure the other team would win).

The pre-registration for Experiment 6, only made predictions regarding cases where advice agrees with advisees' initial opinion. The original analyses plan was to run a linear regression on adviser influence with adviser type (expert vs. novice) and information type (ambiguous vs. unambiguous) as the independent variables. However, after data collection, there were a considerable number of trials where the advice disagreed with participants initial answer in the ambiguous evidence trials. Therefore, the hypothesis was revised (prior to data analyses) for the current and subsequent experiments which included prediction that participants would decrease their confidence in trials where the advisers disagreed with their initial answer. Additionally, data analysis was carried out separately for unambiguous evidence and ambiguous evidence trials. Furthermore, initial agreement was considered in the analyses of ambiguous evidence trials.

For the analysis of ambiguous evidence trials, two separate t-tests were first conducted to test if advisees increase their confidence when an adviser agrees but decrease their confidence when an adviser disagrees (H1). The first t-test examined if overall adviser influence was greater than zero for agreeing trials. The second t-test investigated if adviser influence was negative for disagreeing trials. In order to test if this change in confidence differed by adviser (H1), a 2X2 ANOVA was conducted with adviser influence as the dependent variable and initial agreement (agree vs. disagree) and adviser (expert vs. novice) as the independent variables.

For the unambiguous evidence trials, two t-tests were also conducted. One examined if the adviser influence was below zero (i.e., participants became less confident after receiving advice) and another tested if adviser influence was greater for expert compared to the novice adviser (H2). In addition, a linear mixed model (LMM) was fitted to further examine the data, this analysis was not pre-registered. Participants' adviser influence was included as the dependent variable. The distance between participants' initial confidence and adviser's confidence was calculated as a proxy for the degree of agreement between the advice and participants' initial answer. The confidence distance was calculated as below:

$$\text{Confidence distance} = \text{adviser's confidence} - \text{participant's initial confidence}$$

Confidence distance, adviser (expert vs. novice) type, and the interaction term of these two variables was included in the LMM as fixed effects while participants were included as a random effect.

As an additional measure of adviser influence, weight of advice was calculated. Weight of advice is a measure of advice adoption generally used in estimation tasks (Bailey et al., 2023). There is a possibility that participants would perceive the current task as an estimation task and believe that there is a correct confidence value. Weight of advice was calculated as followed:

$$\text{Weight of advice} = \frac{\text{participant's final confidence} - \text{participant's initial confidence}}{\text{adviser confidence} - \text{participant's initial confidence}}$$

Weight of advice (WOA) measures the degree to which participants move from their own initial confidence estimate to the confidence estimate provided by the adviser. Trials where adviser confidence and participants' initial confidence distance was less than 2 points away (i.e. absolute difference is less than two) were dropped from the analysis as a small difference in participants' initial answer and advice makes the WOA values unreliable (Soll et al., 2022). Additionally, following WOA cut off points from previous studies, values that fell outside of -1 and 2 were excluded from the analysis (Bailey et al., 2023).

In addition to the pre-registered hypotheses, participants' change in the weighting of the four basketball statistics was tested (H3). Change in weighting of the basketball statistics was estimated by first calculating the difference in each measure between the Home Team and the Away Team. For example, if the effective field goal percentage (eFG%) for the Home Team was 57% and 60% for the Away Team, the difference would be 3. Then, the difference was standardised against all collected basketball game data (i.e., standardised using all 168 games collected not solely the 48 games used in experimental trials). The standardised difference was calculated for the four basketball statistics in all trials. After, a linear regression was fitted for each participant, with the participant's initial confidence

as the dependent variable and the four basketball statistics as the independent variable. The linear regression was conducted separately for the first half (first block) and the second half (second block) of the experiment. Therefore, two sets of beta values were derived for each participant. Four t-tests were executed, one for each basketball statistics, to compare the beta values between the first and the second block. The difference in the beta values would indicate a shift in the weighting of the basketball statistics.

### 3.2.2 Results

***Participant Accuracy.*** In the 48 trials, the expert and novice adviser appeared equal number of times (24 trials each). The expert adviser's overall accuracy in the 24 trials was 88% and the novice adviser's accuracy was 92% across the 24 trials. This difference in accuracy was due to the particular subset of trials selected for the experiment, where the novice advisers' predicted winning team was correct in 22 trials (vs. 21 for expert). When accuracy was calculated across all 168 basketball game data used to create each advisers' regression coefficients, the expert adviser was 84% accurate whereas the novice adviser's accuracy was 80%. The higher accuracy of the novice adviser in the experimental trials were unlikely to have influenced participants' perception of adviser expertise because when feedback was provided, the expert was more accurate in their prediction (i.e., training trials) and feedback was not provided during the experimental trials. This was also confirmed by the expert's higher adviser competency rating compared to the novice.

Table 3.2 presents participants' accuracy in predicting the winning team in their initial and final answer (although note that they did not receive feedback in the task), the percentage of trials participants' initial answer and advice agreed, and the percentage of times participants changed their initial answer following advice. In trials with

unambiguous evidence, participants' initial answer mostly agreed with advice. Hence, participants rarely changed their answer. The difference in degree of advice-taking between the two advisers was more apparent in the ambiguous evidence trials. Participants changed their answer more often following the expert's advice,  $t(29) = -6.32$ ,  $p < .001$ . This demonstrates that the expert adviser was more influential.

**Table 3.2** Participant accuracy by trial and adviser type in Experiment 6

	Initial Accuracy (%)	Final Accuracy (%)	Agreeing Trials (%)	Changed Answer (%)
<b>Ambiguous Evidence</b>				
Expert	53	63	63	23
Novice	56	58	63	7
<b>Unambiguous Evidence</b>				
Expert	99	98	99	3
Novice	91	90	99	2

**Adviser Rating.** Participants' perceived competence of each adviser was collected. Expert adviser's rating ( $M_{Exp} = 4.3$ ,  $SE_{Exp} = 0.18$ ) was significantly higher than the novice's ( $M_{Nov} = 2.6$ ,  $SE_{Nov} = 0.23$ ) rating,  $t(29) = 7.77$ ,  $p < .001$ ,  $d = 1.42$ . Thus, the adviser expertise manipulation was successful, and participants perceived the expert adviser to have higher expertise compared to the novice adviser.

**Adviser Influence.**

Out of the 48 trials, 24 were test trials where the advisers displayed high confidence despite ambiguous external evidence or expressed uncertainty despite unambiguous external evidence. The main analyses were conducted with these trials (i.e., excluding filler trials). There were equal number of test trials with the expert and the novice adviser.

Ambiguous evidence trials.

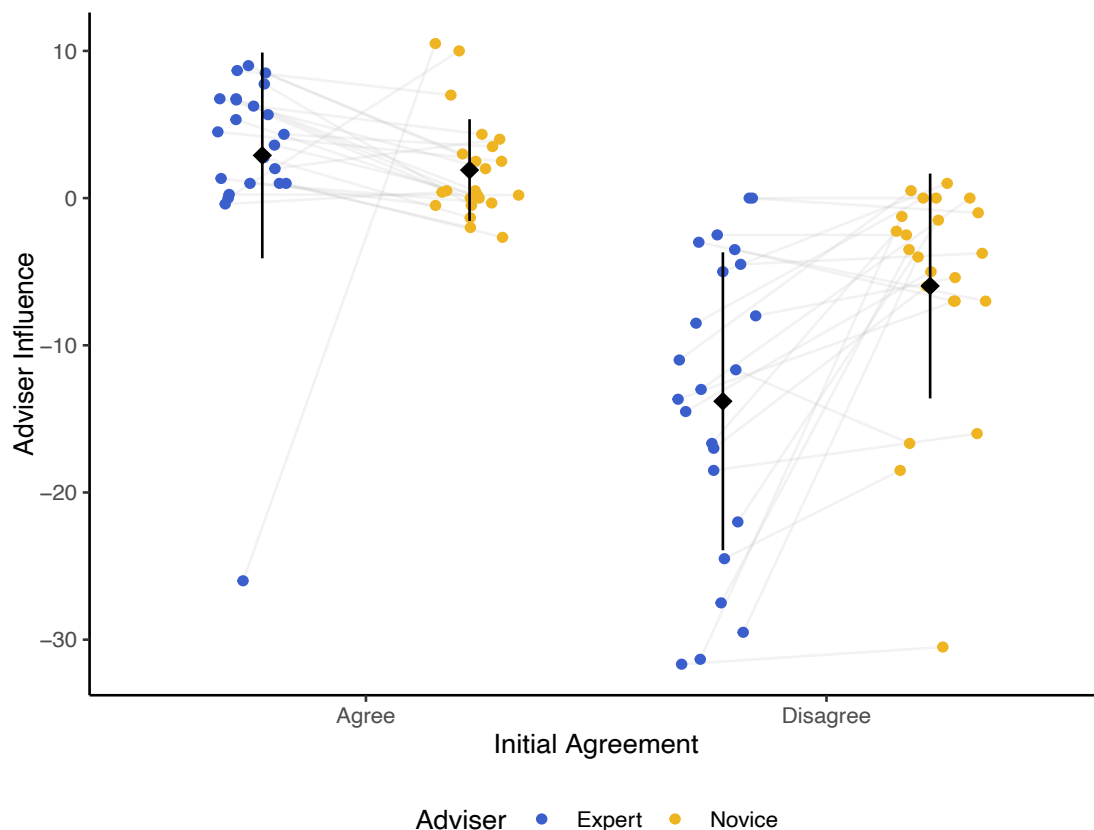
In ambiguous evidence trials the adviser expressed high confidence. H1 predicted that advisers' agreement with the participant's initial prediction would lead to increase in participants' confidence, while disagreement would reduce confidence, with these effects being stronger for the expert adviser compared to the novice adviser. A t-test showed that when an adviser agreed, participants' confidence change (i.e., adviser influence measure) was significantly greater than zero ( $M_{agree} = 2.84$ ,  $SD_{agree} = 3.62$ ),  $t(29) = 4.30$ ,  $p < .001$ ,  $d = 0.79$ . This shows that participants became more confident when an adviser agreed with them. There was a considerable number of trials where advice disagreed with a participant's initial answer. Therefore, in addition to the pre-registered hypothesis, which exclusively focused on agreeing trials, participants' confidence change in disagreeing trials was also investigated through a t-test. This showed that participants generally decreased their confidence when advisers disagreed ( $M_{Disagree} = -10.45$ ,  $SD_{Disagree} = 8.16$ ),  $t(27) = -6.78$ ,  $p < .001$ ,  $d = -1.28$ .

A 2X2 ANOVA was conducted with adviser influence as the dependent variable. Whether participants initial answer and advice agreed (agree vs. disagree), adviser type (expert vs. novice) and their interaction were included as independent variables in the ANOVA. Not all participants had data in all four trial types to run a 2X2 ANOVA. Therefore, these participants had to be removed from the data analysis due to incomplete cases. Hence, 7 participants were removed and a total of 23 participants were included in the analysis. Participants' existing basketball knowledge was included as a covariate.

The analyses revealed a significant main effect of agreement ( $F(1,21) = 56.63$ ,  $p < .001$ ,  $\eta^2 = .451$ ), reflecting that participants generally increased confidence with adviser

agreement and decreased their confidence with disagreement. There was also a significant main effect of adviser type ( $F(1,21) = 7.09, p = .02, \eta^2 = .06$ ) and, critically, a significant interaction between initial agreement and adviser type ( $F(1,21) = 10.09, p = .005, \eta^2 = .10$ ). The interaction effect was further analysed through pairwise comparison with Bonferroni correction. Adviser influence did not differ significantly by adviser for agreeing trials ( $M_{exp} = 2.90, SE_{exp} = 1.46; M_{nov} = 1.90, SE_{nov} = 0.72; p = .601, d = 0.18$ ), although one participant exhibited a significant decrease in their confidence (e.g., change their answer) following agreement from expert adviser (Figure 3.3). However, adviser influence was significantly more negative for trials with the expert adviser ( $M_{exp} = -13.80, SE_{exp} = 2.11$ ) compared to the novice adviser ( $M_{nov} = -5.97, SE_{nov} = 1.59$ ) in disagreeing trials ( $p < .001, d = -0.87$ ). There was a significant main effect of existing basketball knowledge ( $F(1,21) = 6.92, p = .002, \eta^2 = .09$ ). Participants with greater basketball knowledge changed their confidence more.

**Figure 3.3** Adviser influence in ambiguous evidence trials in Experiment 6



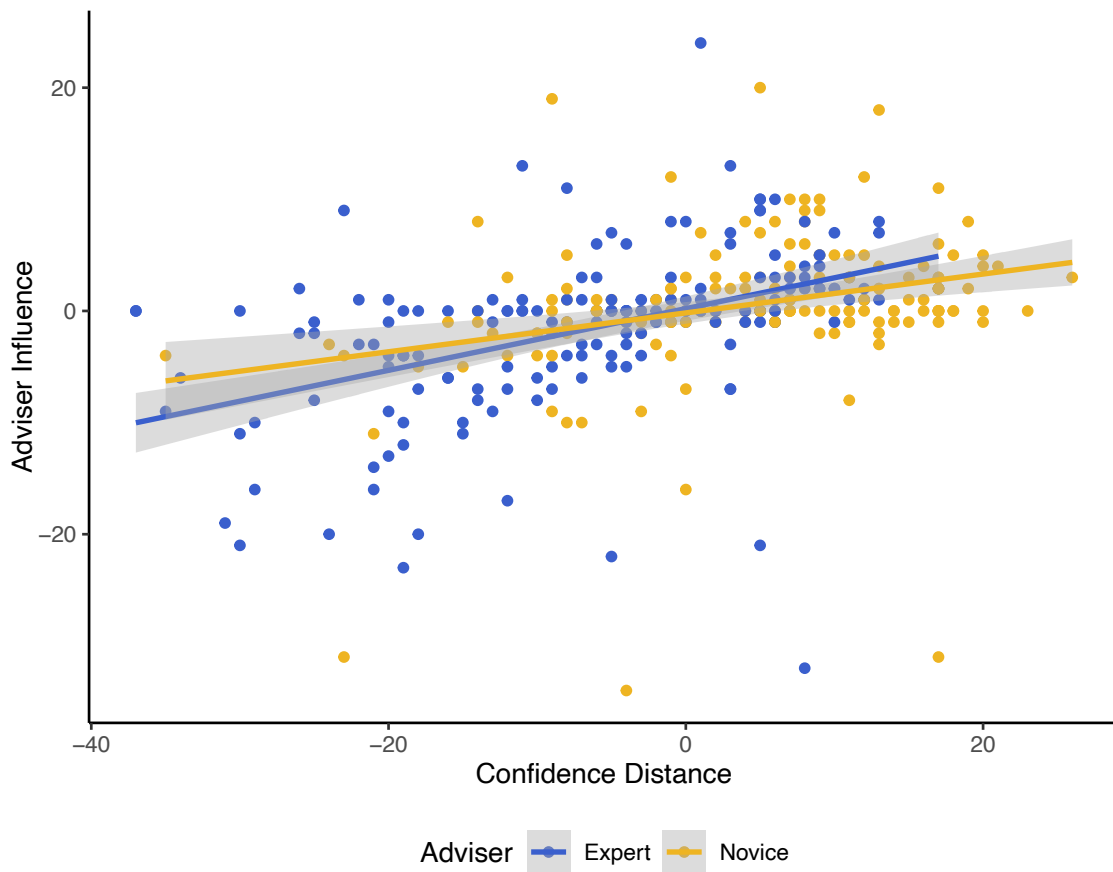
Unambiguous evidence trials.

For trials with unambiguous information, participants' initial answer agreed with advice most of the time (95% of trials), therefore, analyses were limited to these trials. According to H2, when the external evidence is unambiguous, but the adviser is not confident, participants should generally decrease their confidence, and the decrease in confidence should be greater for the expert adviser than the novice adviser. Thus, a t-test was conducted to examine if participants' change in confidence (i.e., adviser influence) was below zero. Although the overall confidence change was negative ( $M = -0.47, SD = 6.88$ ), the value was not significantly below zero,  $t(29) = -0.89, p = .382, d = -0.16$ . In addition, the adviser influence value significant differed between the expert ( $M_{Exp} = -1.56, SD_{Exp} = 3.79$ ) and the novice adviser ( $M_{Nov} = 0.56, SD_{Nov} = 3.03$ ),  $t(29) = -3.47, p = .002, d = -0.63$ . Where only adviser influence of the expert was significantly less than zero,  $t(29) = -2.26, p = .03, d = -0.41$ .

Although this analysis was not pre-registered, a linear mixed model (LMM) was fitted with the adviser influence measure as the dependent variable to further examine the data. Confidence distance (difference between participants' initial answer and advice), adviser (expert vs. novice) type, and the interaction term of these two variables were included as fixed effects while participants was included as a random effect. Participants' existing basketball knowledge was included as a covariate. Analyses found that confidence distance had a significant main effect on adviser influence ( $b = 0.28, SE = 0.03, p < .001$ ). This demonstrates that adviser influence varied depending on confidence distance. More specifically, when the confidence distance was negative, participants tended to decrease their own confidence. In contrast, participants generally increased their confidence post-advice in trials where confidence distance was positive. However, there was no main

effect of adviser type ( $b = 0.43$ ,  $SE = 0.34$ ,  $p = .215$ ) and the interaction effect of confidence distance and adviser was also not significant ( $b = 0.05$ ,  $SE = 0.03$ ,  $p = .086$ ). Although the interaction was not significant, the regression slope was steeper for the expert than for the novice (Figure 3.4). This suggest that participants were slightly more sensitive to confidence distance with the expert. There was no main effect of participants' existing basketball knowledge ( $b = -0.62$ ,  $SE = 0.71$ ,  $p = .387$ ).

**Figure 3.4** *Adviser influence following advice in unambiguous evidence trials in Experiment 6*

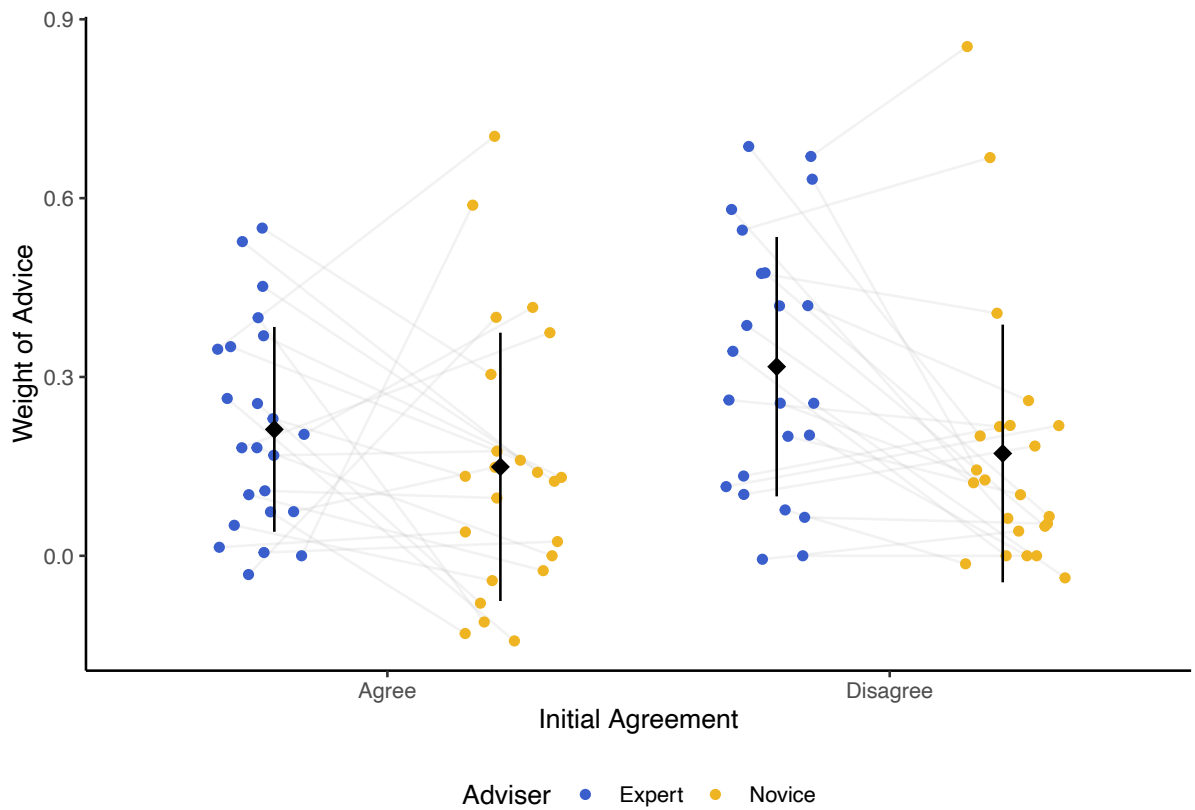


**Weight of Advice.** Weight of advice (WOA) was calculated as an additional measure of advice adoption. WOA is a widely used measure in estimation tasks often used in advice-taking research. Although participants were asked to report their confidence, some participants may have perceived the current experiment as an estimation task. Therefore, the WOA value might better reflect advice adoption for these participants. Additionally,

considering the wide use of WOA, the measure could facilitate comparison between the current experiment and previous literature. Analyses were once again carried out separately for ambiguous evidence and unambiguous evidence trials.

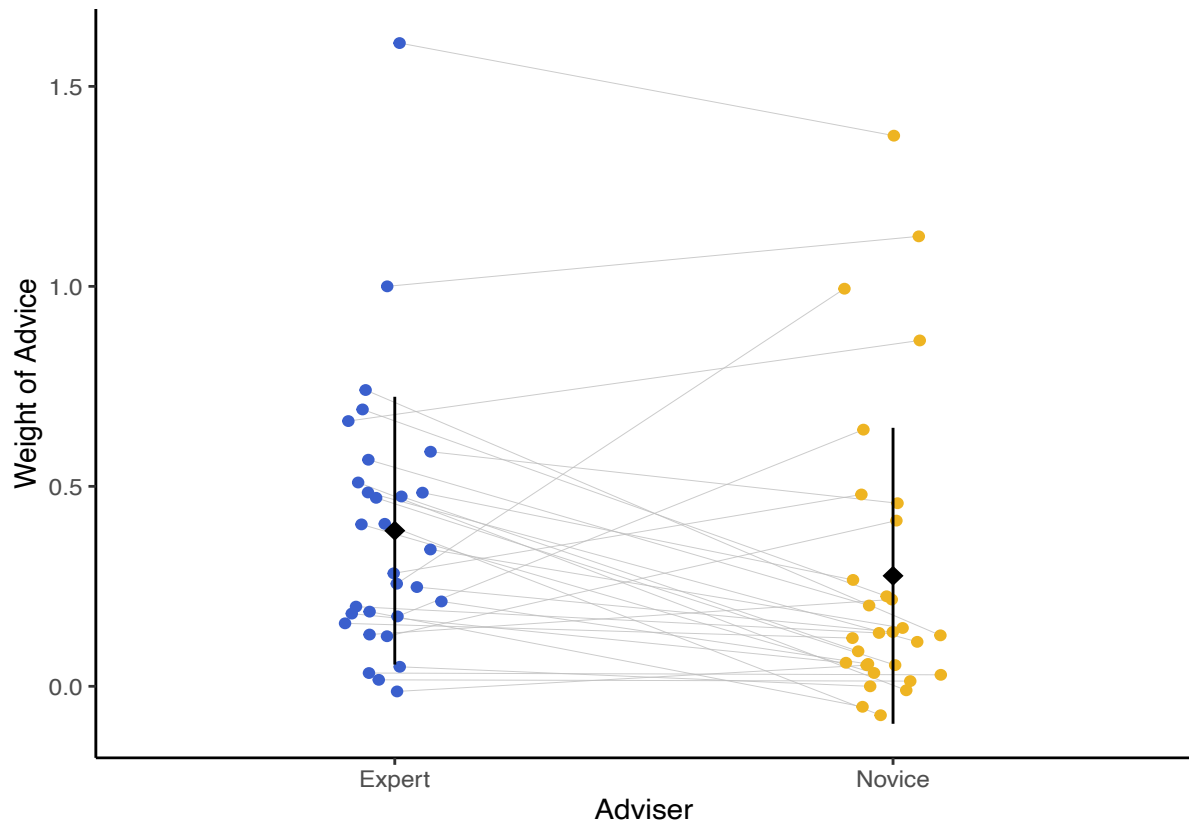
*Ambiguous evidence trials.* A repeated measure 2 (agree vs. disagree) x 2 (expert vs. novice) ANOVA was carried out for the 23 participants who had data in all four trial types with weight of advice (WOA) as the dependent measure. There was a significant main effect of agreement ( $F(1,22) = 4.89, p = .038, \eta^2 = .024$ ), indicating that WOA was greater in the disagreeing trials ( $M_{disagree} = 0.24, SE_{disagree} = 0.03$ ) compared to the agreeing trials ( $M_{agree} = 0.18, SE_{agree} = 0.03$ ). The main effect of adviser type ( $F(1,22) = 9.29, p = .006, \eta^2 = .061$ ) was also significant showing that weight of advice for the expert adviser ( $M_{exp} = 0.27, SE_{exp} = 0.03$ ) was greater than WOA for the novice adviser ( $M_{nov} = 0.16, SE_{nov} = 0.03$ ). The interaction between agreement and adviser type was not significant ( $F(1,22) = 1.12, p = .301, \eta^2 = .010$ ), suggesting that participants' tendency to adopt the expert's advice more than the novice's advice is not contingent on initial agreement.

**Figure 3.5** *Weight of Advice of ambiguous evidence trials in Experiment 6*



*Unambiguous evidence trials.* In the unambiguous evidence trials, participants' initial answers rarely disagreed with advice. Therefore, the weight of advice (WOA) analysis only included agreement trials. A t-test showed that the WOA value did not significantly differ by adviser (expert vs. novice),  $t(29) = 2.04, p = .051, d = 0.37$ . However, the expert adviser's ( $M_{exp} = 0.39, SE_{exp} = 0.06$ ) WOA was numerically higher than the novice adviser's WOA ( $M_{nov} = 0.28, SE_{nov} = 0.07$ ).

**Figure 3.6** *Weight of Advice of unambiguous evidence trials in Experiment 6*



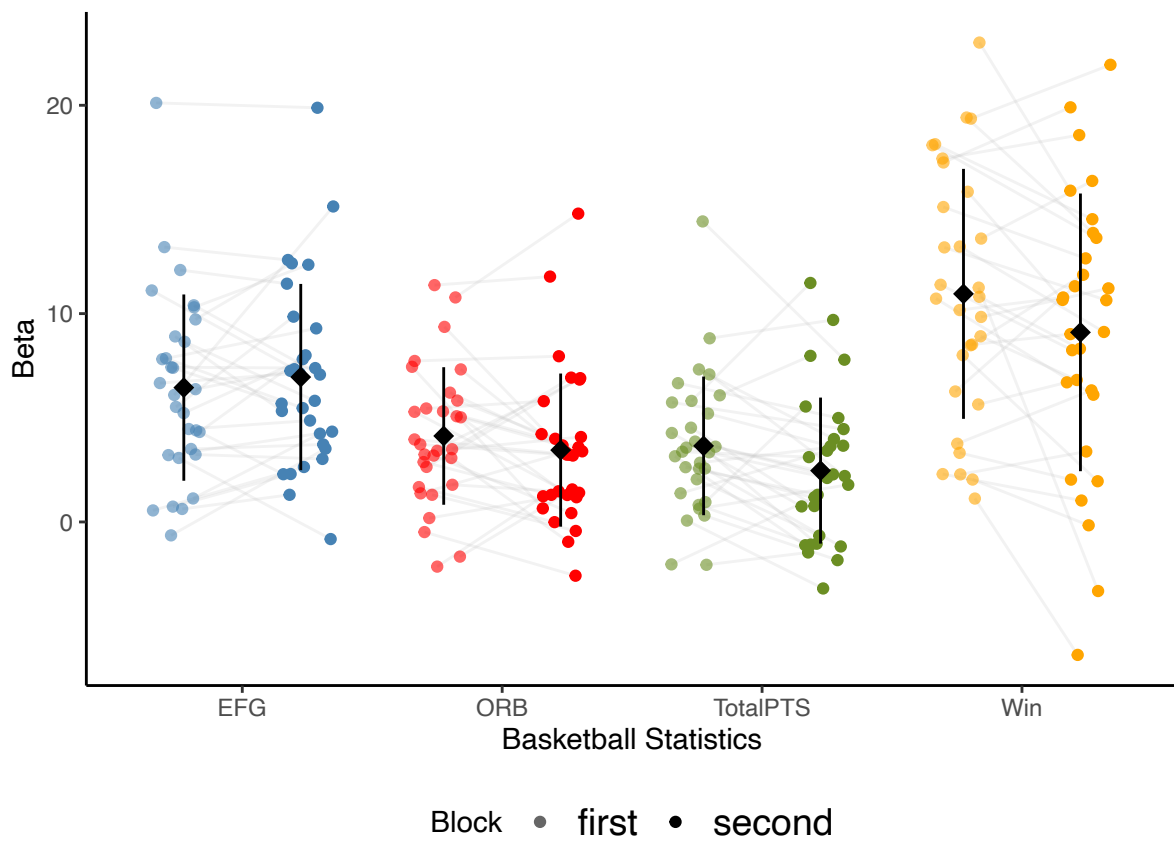
**Advisers' Variable Use Ranking.** Participants also guessed which basketball statistics each adviser prioritised by ranking the four statistics from used the most to the least. Kendall's rank correlation coefficient was calculated to determine participants' accuracy of the estimate. Participants were better able to estimate the expert's ( $r_{expert} = 0.26$ ) weighting of the basketball statistics compared to the novice's ( $r_{novice} = -0.28$ ) weighting,  $t(29) = 4.12, p < .001, d = 0.75$ . The estimate for the expert adviser was significantly different from zero,  $t(29) = 2.60, p = .014, d = 0.48$ . This was also the case for the novice adviser,  $t(29) = -2.68, p = .012, d = -0.49$ . The expert and novice advisers prioritised the four basketball statistics in the same order, but the weighting of each basketball statistic varied. For example, both advisers used eFG% the most in their decision-making but the novice placed less weight. It is possible that participants assumed the two advisers' ranking of the four basketball statistics differed and thus, the correlation coefficient for

the novice is negative. Furthermore, the difference in Kendall's ranking correlation coefficient between the expert and the novice demonstrate that participants did not simply give the same ranking for both advisers but made their predictions deliberately.

**Participants' variable use.** In addition to the pre-registered analyses, it was hypothesised that participants would change their weighting of the four basketball statistics after repeated interaction with the advisers (H3). To test this, participants' use of the four basketball statistics in the first half of the trials and the second half of the trials were compared. In light of the absence of objective feedback, the change in participants' use of the four basketball statistics would indicate that participants were learning and adopting the advisers' use of the basketball statistics from the advisers' advice and associated confidence.

There was a significant decrease in the use of total points scored the previous month from the first block ( $M_{first} = 3.65$ ,  $SE_{first} = 0.61$ ) to the second block ( $M_{second} = 2.47$ ,  $SE_{second} = 0.64$ ),  $t(29) = 2.30$ ,  $p = .030$ ,  $d = 0.42$ . This was also the case for the use of percentage of wins this month ( $M_{first} = 10.95$ ,  $SE_{first} = 1.10$ ;  $M_{second} = 9.10$ ,  $SE_{second} = 1.22$ ;  $t(29) = 2.45$ ,  $p = .020$ ,  $d = 0.45$ ). There was no significant difference in the use of eFG ( $M_{first} = 6.49$ ,  $SE_{first} = 0.82$ ;  $M_{second} = 6.95$ ,  $SE_{second} = 0.82$ ),  $t(29) = -0.97$ ,  $p = .342$ ,  $d = -0.18$  from the first half to the second half of the trials. This was also the case for ORB ( $M_{first} = 4.13$ ,  $SE_{first} = 0.60$ ;  $M_{second} = 3.45$ ,  $SE_{second} = 0.68$ ),  $t(29) = 1.26$ ,  $p = .217$ ,  $d = 0.23$ ). Among the four basketball statistics, the advisers were less reliant on total points scored the previous month and percentage of wins this month. This suggests that participants adopted the advisers decision-making strategy to an extent.

**Figure 3.7** Change in use of the four basketball statistics by block in Experiment 6



In addition to comparing first and second half of the trials, trial effect was also examined. Initially, a linear mixed model was conducted, with participants' initial confidence as the dependent variable and the interaction between predictors (i.e., four basketball statistics) and trial order were included as the fixed effects. Participants were included as the random effect. However, the model found limited variance in participant initial confidence, and thus it was not possible to run a linear mixed model. As an alternative, linear regression with the interaction between the four basketball statistics and trial order was conducted for each participant. The average beta for the interaction between eFG and trial order was not significantly different from zero ( $M_{EFG} = -0.01$ ,  $SE_{EFG} = 0.02$ ),  $t < 1$ ,  $d = 0.07$ . The average beta of the interaction between ORB and trial order was also not significantly different from zero ( $M_{ORB} = -0.02$ ,  $SE_{ORB} = 0.03$ ),  $t < 1$ ,  $d = 0.15$ . On the other hand, the beta coefficient between total points scored the previous month and trial order was significantly different from zero ( $M_{PTS} = -0.06$ ,  $SE_{PTS} = 0.03$ ),  $t(29) = -2.30$ ,  $p = .029$ ,  $d =$

0.42. This was also the case for wins this month ( $M_{WIN} = -0.07$ ,  $SE_{WIN} = 0.03$ ),  $t(29) = -2.11$ ,  $p = .044$ ,  $d = 0.38$ . This shows that participants decreased their use of total points scored the previous month and wins this month as trials passed.

### 3.2.3 Discussion

Experiment 6 explored how people process unexpected adviser confidence in relation to the external information, using a basketball game outcome prediction task. In particular, the experiment investigated how people process confidently given advice despite ambiguous evidence and advice expressing uncertainty even though the external information is unambiguous.

Prior to examining the impact of adviser confidence and advice-taking context, overall perception and advice adoption for the two advisers (expert vs. novice) were examined. Consistent with previous literature (e.g., Bonnacio & Dalal, 2006; Bonnacio & Dalal, 2010; Sniezek et al., 2004; Van Swol & Sniezek, 2005), participants perceived the expert adviser as more competent and were also more likely to utilise their advice. For instance, participants changed their answer more often following the expert's (vs. novice's) advice when the external information was ambiguous.

The main objective of the experiment was to investigate advice-taking when external information ambiguity varied. When the external information was ambiguous, participants tended to increase their confidence after advice that agreed with their initial prediction. However, when the advisers disagreed, participants became less confident in their answer and sometimes even switched their predicted winning team. This finding provides support for H1 and previous literature with similar findings (e.g., Birnbaum &

Stegner, 1979; Harvey & Fischer, 1997). Decrease in confidence was considerably larger after advice from an expert compared to a novice adviser.

When the external information was unambiguous, but the advisers displayed low confidence, participants tended to decrease their confidence, particularly when paired with an expert adviser. This result differs from effects typically observed in previous studies that find agreement from others increases confidence (Julian et al., 1968; McGarty et al., 1993; Schultze et al., 2015; Soll et al., 2021). The results also diverge from a simple Bayesian model which suggests that any form of agreement should increase confidence as it leads to higher probability of being correct, and this increase would be greater with a more trusted (expert) adviser. A possible explanation for this effect is that the advisers displaying low confidence may have led people to doubt their assessment of the external information. When the participants were overly confident compared to the advisers, they would re-evaluate their use of the four basketball statistics and their overall decision-making strategy. In line with this, there was some change participants' use of the basketball statistics over time even though no explicit feedback was provided. In the second half of the trials, participants decreased their use of total points scored last month and percentage of wins this month, which were statistics used the least by the advisers. This possibly indicates that participants deduced the advisers' weighting of the basketball statistics and adopted the advisers' decision-making strategy.

On the other hand, an alternative explanation of the findings is that people treat the task as an estimation task where there is a *correct* confidence value. Therefore, people might have been simply lowering their confidence to match the adviser's confidence rather than subjectively feeling less confident about their answer and decision-making strategy. To

address this issue, two follow-up studies were conducted where either the advisers' confidence or the participants' confidence was recorded using linguistic expressions. Linguistic expressions tend to represent a range of confidence that may slightly vary for each individual (Hamm, 1991). Therefore, we believed that advisers' use of the linguistic expressions (instead of precise numeric confidence values) would limit the tendency to copy or match the advisers' confidence. Thus, the follow-up experiments would better be able to capture genuine change in confidence.

### **3.3 Experiment 7**

Experiment 6 showed that participants decreased their confidence even when an adviser agreed with their initial answer when the external information is ambiguous. However, it was unclear if this change in confidence reflected a genuine change in subjective confidence or participants' tendency to match the advisers' confidence. In other words, participants could have treated the task as an estimation task and believed that there was a "correct confidence" value. Therefore, in Experiment 7, advisers' confidence was provided using linguistic expressions. Linguistic expressions do not have a specific point confidence value. Thus, participants are less likely to believe there is a correct single value and it would also be difficult for participants to exactly copy the advisers' confidence. Five linguistic expressions of confidence were adopted from previous literature.

#### **3.3.1 Methods**

**Participants.** In Experiment 7, 60 participants (29 = Female, 31 = Male,  $M_{\text{age}} = 37.0$   $SD_{\text{age}} = 12.3$ ) were recruited through Prolific ([www.prolific.com](http://www.prolific.com)). The number of participants was increased as the range of adviser confidence reduced to only five levels. Therefore,

participants were less likely to change their confidence unless the advisers' confidence substantially differed from their own confidence, thus reducing the overall adviser influence. Hence, increasing the number of participants would help better detect the change in participant confidence in response to advice. Informed consent was obtained from all participants before commencing the experiment and participants were compensated £9/hour based on the median duration of the experiment.

**Task.** The task in Experiment 7 was identical to Experiment 6. Participants predicted the outcome of basketball games based on four statistical information: effective field goal rate (eFG%), offensive rebound rate (ORB%), points scored the previous month, and percentage of wins this month. Participants chose between Home Team and Away Team by moving a slider on a horizontal scale ranging from 100% Home Team to 100% Away Team. The mid-point (50% confident) was not selectable thus, it was mandatory to answer the questions.

**Advisers.** Consistent with Experiment 6, there was one expert and one novice adviser. Each advisers' prediction of the basketball game outcome was constructed in the same way as the previous study. However, in this study, the advisers gave their confidence level using linguistic expressions rather than numerical values. The linguistic expressions were adopted from Hamm (1991). Hamm (1991) constructed 19 lists of verbal expressions and asked participants to assign a numerical value to each expression. Among these, five linguistic expressions of confidence where most participants assigned a similar numeric range was selected. In the current experiment, each linguistic expression represented approximately a 10 percent range of confidence. More specifically, the expressions were *Toss-up* (50.1%-59.9% confidence), *Better than even chance* (60-69.9%), *Quite Likely*

(70-79.9%), *Highly Probable* (80-89.9%), and *Almost Certain* (90-100%). For example, if the adviser was 93% confident that the Home Team would win, he would now say “I am almost certain that Home Team would win”. Both advisers (expert vs. novice) used the same set of linguistic expressions of confidence.

***Procedure.*** Experiment 7 followed the same procedure as Experiment 6. After reading the task instructions and learning about the statistics, participants completed a training block that consisted of the same three trials from Experiment 6. The advisers agreed on the (correct) answer in the first two trials, in the third training trial the expert and novice predicted different team would win but gave identical confidence level (unlike Experiment 6 where the expert was more confident). This was because the adviser confidence numerically differed but fell in the same confidence range for the linguistic expression.

There were 48 experimental trials, split into two blocks with the option to rest after the first 24 trials. The test trials were identical to Experiment 6, except that adviser confidence was translated to linguistic expressions following the method outlined above. Upon completion of the test blocks, participants gave adviser competence ratings. Then, participants were asked to guess the four basketball statistics ranking for each adviser (rank from used the most to the least). Lastly, demographic information (age and gender) was collected. Participants' basketball knowledge was initially collected as a covariate in Experiment 6. However, most participants reported having limited knowledge in basketball (None,  $N=12$ ; Poor,  $N = 14$ ; Decent,  $N = 2$ ; Good,  $N = 2$ ). Therefore, it was difficult to interpret the analyses results meaningfully. Thus, this measure was dropped in the current experiment and Experiment 8.

### ***Data Analysis.***

Experiment 6 showed that in the ambiguous evidence trials, participants' initial answer and advice disagreed for a considerable number of trials. Therefore, H1 and data analysis was revised in the pre-registration for Experiment 7. Data analysis for Experiment 7 followed the same method outlined in the Data Analysis section of Experiment 6, but without the covariate (basketball knowledge). Another difference was how the confidence distance (adviser's confidence – participant's initial confidence) measure and the weight of advice (WOA) was calculated. In this experiment, the advisers used linguistic expressions to convey their confidence, which represented a range of values. For both measures, the adviser's confidence needs to be a single numerical value. Therefore, the linguistic expressions were translated to a single numerical value by taking the mid-point of the confidence range for each linguistic expression. For example, *Almost Certain* represented confidence between 90 to 100%. Therefore, confidence of *Almost Certain* was computed as 95% to derive the confidence distance and WOA value.

### **3.3.2 Results**

***Participant Accuracy.*** Similar to Experiment 6, participants' initial answer and advice mostly agreed in the unambiguous evidence trials and thus, participants did not change their initial answer (Table 3.3). On the other hand, participants changed their initial answer more often following the expert's advice in the ambiguous evidence trials,  $t(59) = -4.64, p < .001, d = 0.60$ . This was also reflected in the greater increase in accuracy from initial to final answer trials with the expert's advice,  $t(59) = 2.73, p = .008, d = 0.35$ .

**Table 3.3** Participant accuracy by trial and adviser type in Experiment 7

	Initial Accuracy (%)	Final Accuracy (%)	Agreeing Trials (%)	Changed Answer (%)
Ambiguous Evidence				
Expert	49	57	58	16
Novice	54	58	61	8
Unambiguous Evidence				
Expert	99	99	99	0
Novice	91	91	99	0

**Adviser Rating.** Adviser ratings were collected to measure participants' perceived competence of the expert and the novice advisers. The expert adviser's ( $M_{exp} = 4.5$ ,  $SE_{exp} = 0.13$ ) rating was significantly higher than the novice adviser's ( $M_{nov} = 2.9$ ,  $SE_{nov} = 0.14$ ) rating,  $t(59) = 10.01$ ,  $p < .001$ ,  $d = 1.29$ . This demonstrates that people indeed perceived the expert adviser to be more competent and manipulation of adviser expertise was successful.

**Adviser Influence.** The main measure of interest was how much participants update their initial confidence in response to each adviser's advice. This was measured as adviser influence (change in confidence from initial to final answer), which was calculated in the same manner as Experiment 6. Analyses were carried out separately for the 12 ambiguous evidence trials and 12 unambiguous evidence trials.

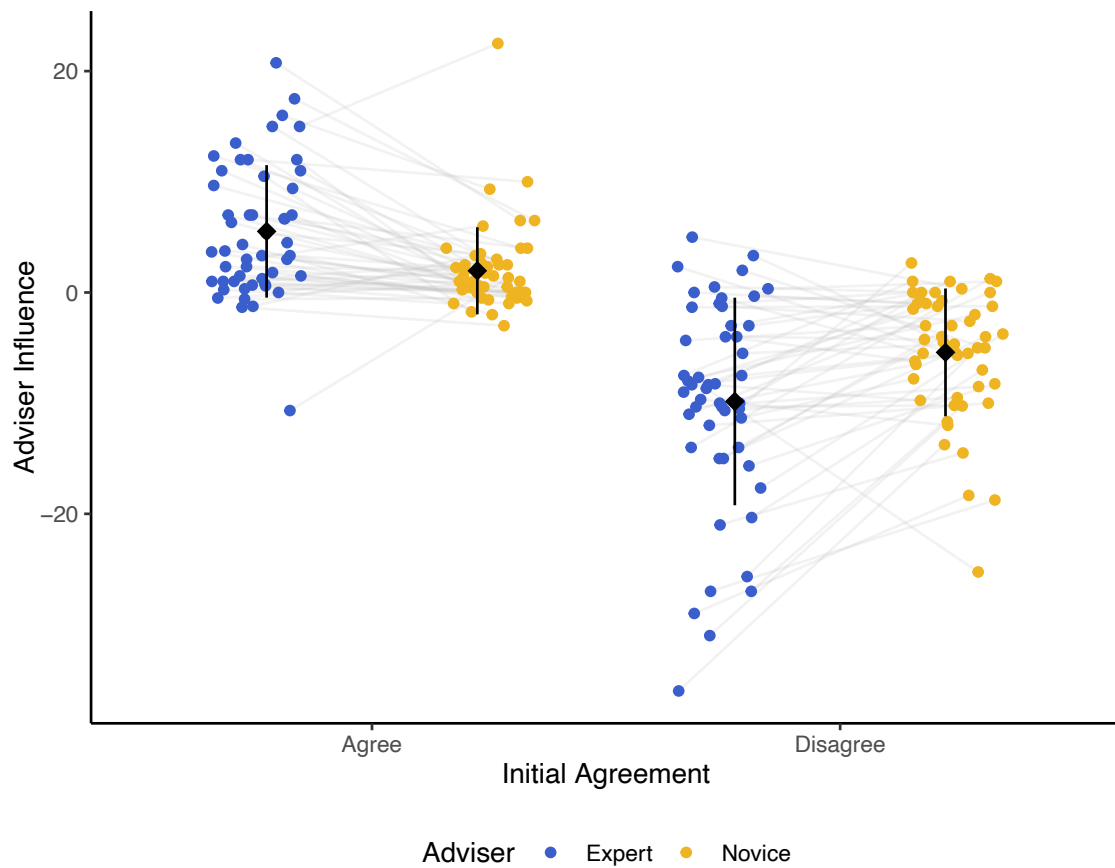
**Ambiguous evidence trials.** For the ambiguous evidence trials, it was hypothesised that agreeing advice would increase advisees' (i.e., participant) confidence while disagreeing advice lowers confidence, with this effect being stronger for the expert adviser (H1). In order to test H1, two t-tests were carried out to examine adviser influence following advice that agrees or disagrees with participants' initial answer. As hypothesised, a t-test

found that adviser influence (participant's final – initial confidence) was positive for trials where participants' initial answer and advice agreed ( $M_{Agree} = 4.01$ ,  $SD_{Agree} = 4.38$ ),  $t(59) = 7.09$ ,  $p < .001$ ,  $d = 0.91$ . On the other hand, adviser influence was negative when the adviser disagreed with the participants' initial answer ( $M_{Disagree} = -7.18$ ,  $SD_{Disagree} = 6.19$ ),  $t(58) = -8.91$ ,  $p < .001$ ,  $d = -1.16$ . This shows that participants generally became more confident when an adviser agreed but decreased their confidence when an adviser disagreed.

A 2 (agree vs. disagree) x 2 (expert vs. novice) repeated measures ANOVA was carried out to examine the adviser influence measure in trials with ambiguous basketball statistics. Nine participants were excluded from the analyses due to not having data in all four trial types (i.e., incomplete cases) and thus, data from a total of 51 participants were analysed. There was a significant main effect of initial agreement ( $F(1,50) = 68.99$ ,  $p < .001$ ,  $\eta^2 = .43$ ). The main effect of adviser type was not significant ( $F(1,50) = 0.45$ ,  $p = .505$ ,  $\eta^2 = .001$ ). However, there was a significant interaction effect between initial agreement and adviser type ( $F(1,50) = 34.78$ ,  $p < .001$ ,  $\eta^2 = .09$ ). Therefore, pairwise comparison with Bonferroni correction was conducted. In the agreeing trials, there was a significant difference in adviser influence by adviser type ( $p < .001$ ,  $d = 0.70$ ), indicating that participants became more confident following the expert's advice ( $M_{exp} = 5.52$ ,  $SE_{exp} = 0.84$ ) compared to the novice's advice ( $M_{nov} = 1.96$ ,  $SE_{nov} = 0.55$ ). There was also a significant difference in adviser influence by adviser type in disagreeing trials ( $p < .001$ ,  $d = -0.57$ ). Adviser influence was significantly more negative following the expert's advice ( $M_{exp} = -9.85$ ,  $SE_{exp} = 1.31$ ) compared to the novice's advice ( $M_{nov} = -5.41$ ,  $SE_{nov} = 0.81$ ). This demonstrates that participants reduced their confidence more after disagreement from the expert. The findings from the t-test and ANOVA taken together supports H1.

**Figure 3.8** Adviser influence following advice in ambiguous evidence trials in Experiment

7



Unambiguous evidence trials. In the unambiguous evidence trials, it was predicted that unexpected low adviser's confidence would lead participants to decrease their confidence despite agreeing advice. This effect was expected to be stronger for the expert adviser (H2). There were only a few trials where participants' initial answer disagreed with the advice. Therefore, these trials were dropped from the analyses. A t-test was first conducted to check if the difference between participants' initial and post-advice confidence (i.e., adviser influence) was negative following advice. However, adviser influence was significantly positive ( $M = 1.98$ ,  $SD = 3.64$ ),  $t(59) = 4.22$ ,  $p < .001$ ,  $d = 0.54$ . This contrasted with Experiment 6 results, where adviser influence was negative following advice in trials with unambiguous evidence. The increase in average confidence was numerically greater for the expert adviser ( $M_{Exp} = 2.03$ ,  $SD_{Exp} = 4.96$ ) compared to

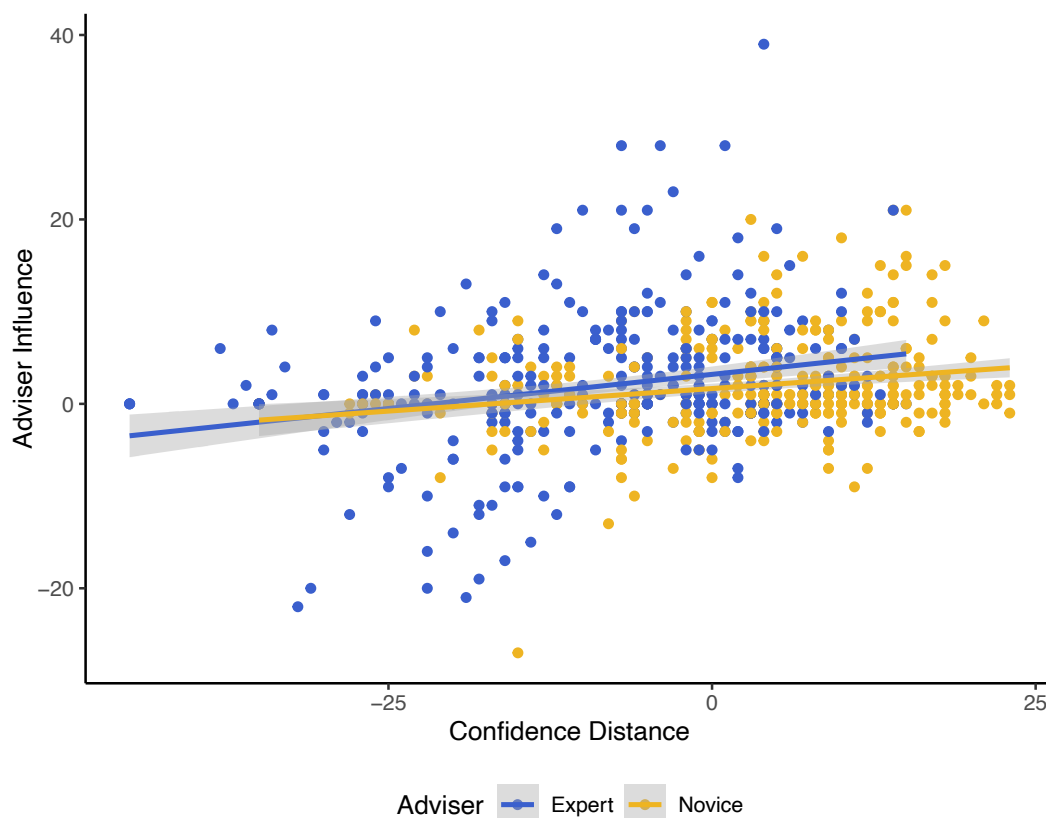
the novice adviser ( $M_{Nov} = 1.89$ ,  $SD_{Nov} = 3.36$ ) but the difference was not significant,  $t(59) = 0.26$ ,  $p = .602$ ,  $d = 0.03$ .

Once again to better understand the trends in the data, a linear mixed model was fitted. Adviser influence was the dependent variable, while adviser type (expert vs. novice) and confidence distance (adviser's confidence – participants' initial confidence) and their interaction term were the fixed effects. Participants were included as a random effect. There was a significant main effect of confidence distance ( $b = 0.20$ ,  $SE = 0.02$ ,  $p < .001$ ), where a more positive confidence distance led to a higher adviser influence while negative adviser influence led to decrease in confidence (i.e., negative confidence difference). The main effect of adviser type was also significant ( $b = 1.17$ ,  $SE = 0.21$ ,  $p < .001$ ), indicating that change in confidence was greater following the expert's advice compared to the novice's advice. The interaction term was also significant ( $b = 0.03$ ,  $SE = 0.02$ ,  $p = .049$ ). This demonstrates that the effect of change in confidence distance on adviser influence was greater with the expert adviser than with the novice adviser. In other words, participants were more sensitive to confidence distance with the expert compared to the novice.

The analysis results and Figure 3.9 show that participants generally increased their confidence following agreement from the advisers. The overall increase in confidence was greater (although not significantly) when an expert agreed. On the other hand, although less so than Experiment 6, there were also cases where participants decreased their confidence in spite of agreeing advice. Decrease in confidence was particularly prominent when the adviser's confidence was much lower (about 20%) than their own confidence and following the expert's advice (Figure 3.9). In the current experiment, the

adviser confidence was conveyed using linguistic expressions which represented a range of confidence. Therefore, participants might have overlooked small differences between their own and the adviser's confidence. Hence, participants may have only decreased their confidence when an adviser expressed substantially lower confidence than themselves. Furthermore, participants still decreased their confidence despite the lack of specific numerical value (as advisers' linguistic expressions represented a range of values) to practice confidence matching. Thus, the findings in the current experiment reduce the likelihood that participants were simply matching their confidence to the adviser's confidence.

**Figure 3.9** *Adviser influence following advice in unambiguous evidence trials in Experiment 7*

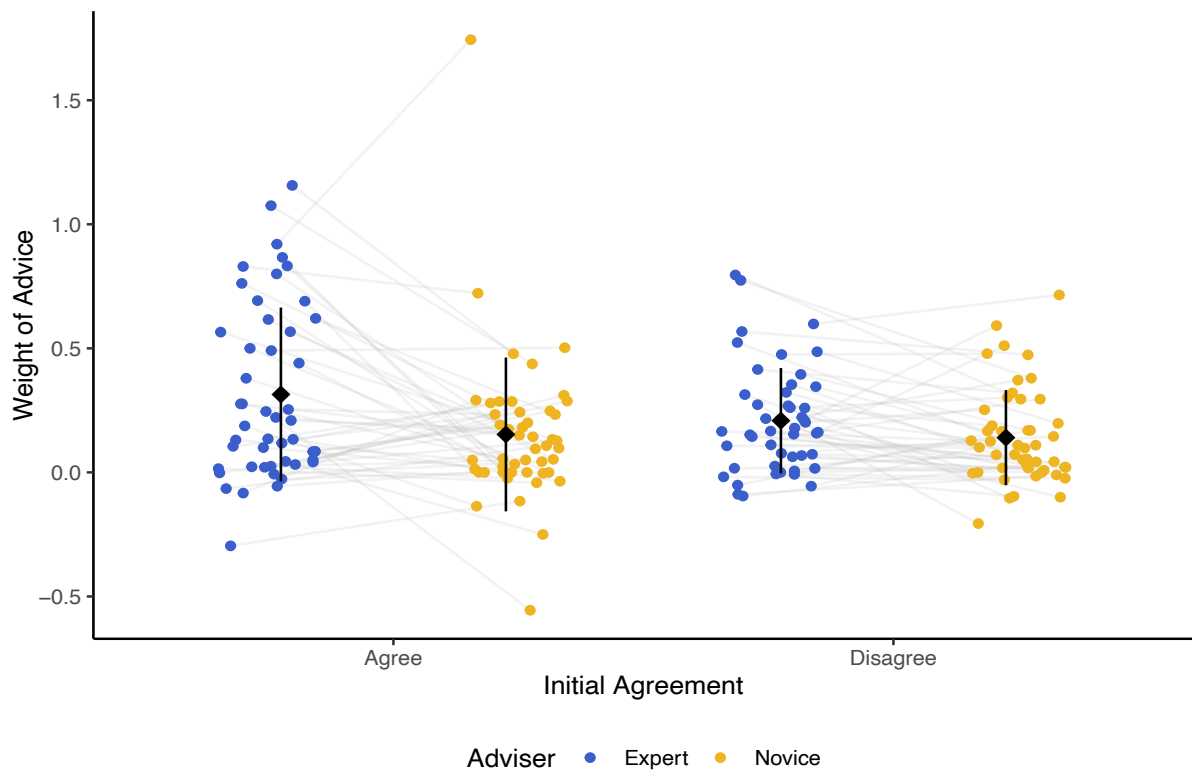


**Weight of Advice.** Similar to Experiment 6, weight of advice (WOA) measure was calculated as an additional measure of advice adoption. In this experiment, the advisers

communicated their confidence through linguistic expressions. Therefore, WOA was calculated by taking the mid-point of the confidence range for each expression, following the method outlined previously. Apart from this difference, WOA was calculated following the same equation outlined in Experiment 6. Weight of advice values where participants' initial confidence and adviser confidence differed by less than 2 points and WOA values less than -1 or greater than 2 were excluded from the analyses. Analyses were carried out separately for ambiguous and unambiguous evidence trials. Prior to presenting the analyses results, it is worth noting that WOA is normally used for estimation tasks. It was unlikely that participants treated the current experiment as an estimation task (considering the advisers' use of linguistic expressions for confidence instead of numerical value). However, the WOA value was still calculated for consistency with Experiment 6 and to allow some comparison with previous literature.

*Ambiguous evidence trials.* The WOA values of ambiguous evidence trials were analysed using a repeated measures 2 (agree vs. disagree) x 2 (expert vs. novice) ANOVA. There were 48 participants remaining after eliminating participants with incomplete cases. There was no significant main effect of agreement ( $F(1,47) = 2.57, p = .116, \eta^2 = .012$ ). This contrasted with Experiment 6, which found that WOA values were greater in disagreeing trials. However, there was a significant main effect of adviser type ( $F(1,47) = 13.80, p < .001, \eta^2 = .04$ ). The expert's WOA ( $M_{exp} = 0.26, SE_{exp} = 0.03$ ) was greater than the novice's WOA ( $M_{nov} = 0.15, SE_{nov} = 0.03$ ). This once again demonstrates that the expert adviser was more influential. Consistent with Experiment 6, the interaction term was not significant ( $F(1,47) = 2.51, p = .120, \eta^2 = .007$ ).

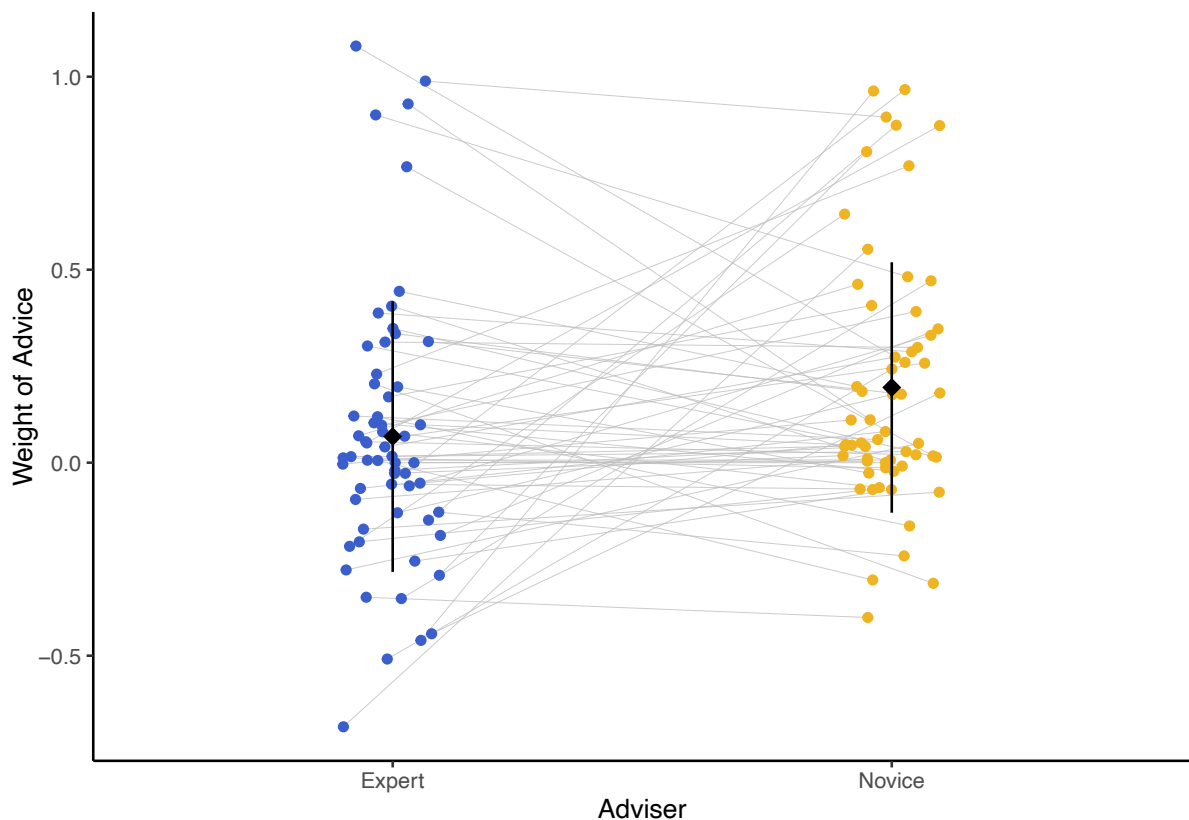
**Figure 3.10** *Weight of Advice of ambiguous evidence trials in Experiment 7*



*Unambiguous evidence trials.* A t-test comparing weight of advice (WOA) for expert and novice was carried out. In contrast to Experiment 6 and the analysis above of ambiguous evidence trials, here the WOA for the expert ( $M_{exp} = 0.07$ ,  $SE_{exp} = 0.04$ ) was significantly lower than the novice's WOA ( $M_{nov} = 0.19$ ,  $SE_{nov} = 0.04$ ),  $t(59) = -2.07$ ,  $p = .043$ ,  $d = -0.27$ . Additionally, the overall WOA values in the current experiment were numerically lower than Experiment 6 ( $M_{exp} = 0.4$ ;  $M_{nov} = 0.28$ ). The difference between the two experiments is likely due to the greater number of negative WOA values in the current experiment (Figure 3.11). There were more negative WOA values following the expert's advice (Figure 3.11), which is likely to have resulted in the expert's overall lower WOA compared to the novice. The negative WOA values are likely to be due to participants becoming more confident than the adviser following agreeing advice. For example, imagine a participant initially predicted that Home Team would win with 70% confidence and an adviser agrees with 60% confidence. If the participant increases their confidence

to 80%, the WOA value would be -1. In Experiment 7, adviser confidence was given using linguistic expressions, which incorporates a range of confidence values. Thus, participants might have not realised that they were overconfident compared to the adviser and became more confident in their judgement. This is also evident in Figure 3.9 where there are considerable number of positive adviser influence values despite negative confidence distance. These analyses demonstrate the limits of extending the standard use of WOA to the present type of tasks where advice and participants' answers are collected using different scales. Therefore, comparison of WOA values in the current experiment and previous literature should be made cautiously.

**Figure 3.12** *Weight of Advice of unambiguous evidence trials in Experiment 7*

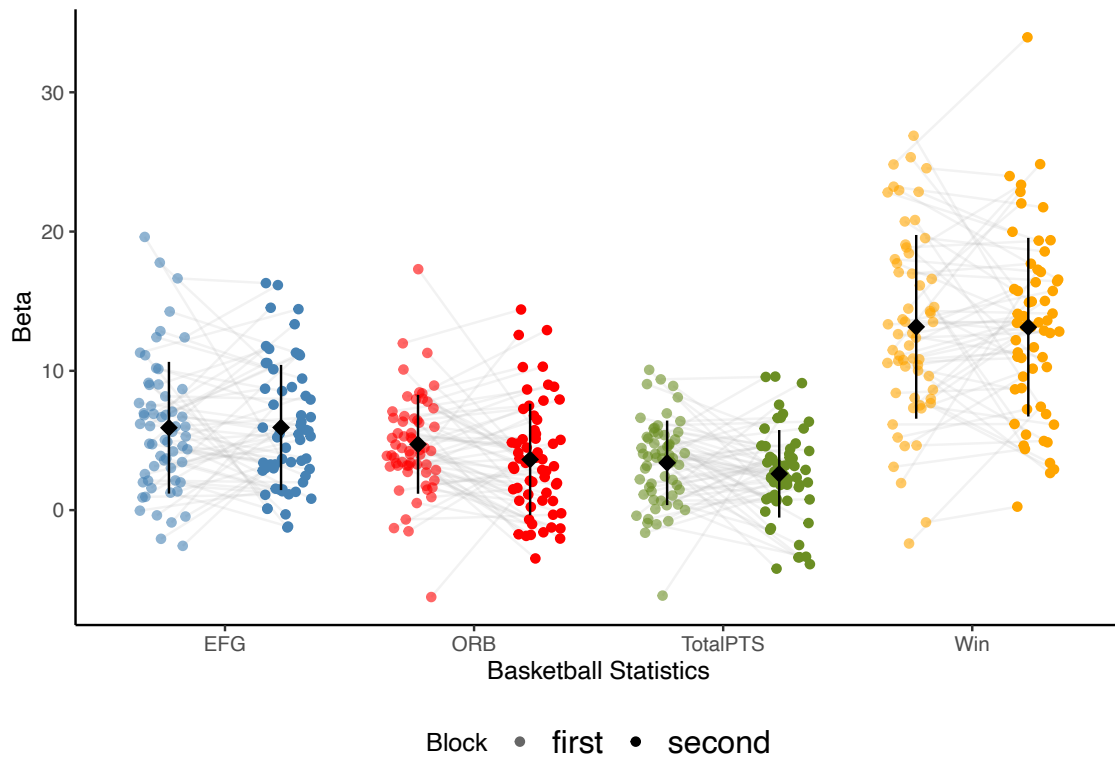


**Advisers' Variable Use Ranking.** Participants were asked to estimate each adviser's use of the four basketball statistics and rank them from used the most to the least. This was to examine if participants were able to deduce the adviser's weighting of the four

basketball statistics. A Kendall's rank correlation coefficient was calculated between participants' predicted ranking and each adviser's actual ranking. Participants were significantly better at ranking the expert's use of the four basketball statistics ( $M_{exp} = 0.16$ ,  $SE_{exp} = 0.07$ ) than ranking the novice's use ( $M_{nov} = -0.33$ ,  $SE_{nov} = 0.06$ ),  $t(59) = 5.19$ ,  $p < .001$ ,  $d = 0.67$ .

**Participants' variable use.** The degree to which each participants used the basketball statistics in their decision-making was estimated for the first and second half of the trials. T-tests were carried out for each of the four basketball statistics, which found no significant difference: effective field goal ( $p = .978$ ,  $d = -0.004$ ), offensive rebound ( $p = .050$ ,  $d = 0.26$ ), total points scored the previous month ( $p = .142$ ,  $d = 0.19$ ) and percentage of wins this month ( $p = .972$ ,  $d = 0.005$ ). However, there was a numerical reduction in the use of offensive rebound and total points scored the previous month. The trial effect was also examined in Experiment 7, following the same method as Experiment 6. The analysis yielded the similar results, where the average beta for the interaction between eFG and trial order was not significantly different from zero ( $M_{EFG} = 0.02$ ,  $SE_{EFG} = 0.02$ ),  $t < 1$ ,  $d = 0.10$ . The average beta of the interaction between ORB and trial order was significantly different from zero ( $M_{ORB} = -0.05$ ,  $SE_{ORB} = 0.02$ ),  $t(59) = -2.27$ ,  $p = .027$ ,  $d = 0.29$ , suggesting participants generally decreased the use of ORB as trials passed. On the other hand, the beta coefficient between total points scored the previous month and trial order was not significantly different from zero ( $M_{PTS} = -0.03$ ,  $SE_{PTS} = 0.02$ ),  $t(59) = -1.43$ ,  $p = .157$ ,  $d = 0.19$ . This was also the case for wins this month ( $M_{WIN} = -0.02$ ,  $SE_{WIN} = 0.03$ ),  $t < 1$ ,  $d = 0.11$ .

**Figure 3.13** *Change in the use of four basketball statistics by block in Experiment 7*



### 3.3.3 Discussion

Experiment 7 was carried out to examine if change in participants' confidence is contingent on the desire to match one's own confidence to the adviser's confidence. In order to test this, the experiment used five linguistic expressions selected from Hamm (1991). Linguistic expressions do not have a specific value but are associated with a range of confidence values. Furthermore, advisers and participants express their confidence using different methods. Therefore, change in participants' confidence is more likely to indicate an actual change in subjective feeling of confidence rather than a result of confidence matching.

Focusing on trials where the basketball statistics were ambiguous, agreement from the advisers led to increase in confidence and disagreement resulted in confidence reduction. This effect was stronger following the expert's advice than the novice's advice.

Additionally, the weight of advice (WOA) value was also greater for the expert adviser. This demonstrates that the expert adviser was more influential, which replicates Experiment 6 results. On the other hand, there were some differences in the results for the unambiguous evidence trials compared to Experiment 6. In contrast to Experiment 6, where the overall adviser influence was negative, participants' overall change in confidence was positive. Furthermore, participants on average increased their confidence more after receiving the expert's advice. Interestingly, the (WOA) value was numerically higher for the novice than the expert adviser. The lower WOA value for expert was due to the greater cases of negative WOA. As adviser confidence was conveyed with five linguistic expressions, participants might have perceived advisers to have similar levels of confidence as themselves even when the adviser was less confident and thus, increased their own confidence, which resulted in negative WOA value. Greater frequency of negative WOA value may also explain the generally lower WOA value in Experiment 7 compared to Experiment 6.

In the case of unambiguous evidence trials, although the overall adviser influence value was positive, this was numerically lower than the overall adviser influence value under agreement in the ambiguous evidence trials. This suggests that participants were less likely to increase their confidence despite agreement from the adviser in the unambiguous evidence trials. Additionally, in some trials, participants decreased their confidence, and this was particularly evident when the advisers were considerably less confident than the participants. In Experiment 7, the advisers used five linguistic expressions of confidence, where each expression is associated with a range of confidence values. Thus, the difference between participant's own confidence and the adviser's confidence was less apparent compared to Experiment 6. Participants might have perceived advisers to be

similarly confident as themselves or viewed the difference as negligible unless the adviser's confidence significantly diverged from their own confidence. Therefore, participants decreased their confidence despite agreement only when the advisers were substantially less confident than themselves. Thus, negative adviser influence values (i.e., participants decrease their confidence) occurred less in Experiment 7 than in Experiment 6. This is likely to have driven the positive adviser influence value in Experiment 7. Even though the effect was smaller than Experiment 6, decrease in confidence despite agreement still occurred which deviates from previous findings that generally shows that agreement from advisers lead to increase in confidence (Julian et al., 1968; McGarty et al., 1993; Schultze et al., 2015; Soll et al., 2022).

Furthermore, the results of the current experiment suggest that people's reduction in confidence despite agreeing advice is not merely because people perceive the task as an estimation task where there is a "correct" confidence value and practice confidence matching (e.g., Friedmann et al., 2024). The linguistic expressions used by the advisers incorporates a range of confidence and therefore it is difficult to deduce a specific numeric confidence. Thus, people are likely to have sensed the advisers' uncertainty and also question their confidence.

However, in contrast to Experiment 6 where participants changed their decision-making strategy there were only small changes in participants' use of the four basketball statistics from first to second half of the trials. For example, there was a non-significant numerical reduction in the use of Total Points Scored. This is the statistic weighted the least by the advisers, suggesting some learning of advisers' decision-making strategy. However, participants also used ORB less in their decision, which is a variable the advisers prioritise.

The participants' limited adoption of the advisers' use of the four basketball statistics compared to Experiment 6 might be due to two reasons. Firstly, the adviser only used five linguistic expressions. Therefore, it might have been more challenging to deduce the advisers' weighting of the basketball statistics. Second, participants might have doubted their decision-making strategy less compared to Experiment 6, which is reflected in greater increase in confidence in unambiguous evidence trials. Therefore, participants did not feel the need to adopt the advisers' weighting of the basketball statistics.

### **3.4 Experiment 8**

In Experiment 8, the advisers' confidence was conveyed using numerical values (same as Experiment 6), but participants were asked to give their answer using linguistic expressions of confidence. The current experiment aimed to investigate if people change their confidence even when using a different response mechanism – linguistic expressions. Another purpose was to ensure that people are interpreting, understanding and using the linguistic expressions of confidence as intended in Experiment 7 and previous literature.

#### **3.4.1 Methods**

**Participants.** Sixty participants (29 = Female, 31 = Male,  $M_{\text{age}} = 37.8$ ,  $SD_{\text{age}} = 12.04$ ) were recruited through Prolific (<https://www.prolific.com>). The number of participants remained larger than Experiment 6 because measure of participant confidence reduced to five levels. As there are only five levels, participants are only likely report a different confidence level following a large change in confidence and a small confidence change would not be detected. Therefore, sample size was increased to compensate for this reduced power to detect smaller effects. Informed consent was obtained from all

participants prior to commencing the study. Compensation rate was £9/hour based on the median duration of the experiment.

**Task.** The task was once again to predict basketball game outcomes using four basketball statistics given for each team. The basketball statistics given were identical to the previous experiments: effective field goal rate (eFG%), offensive rebound rate (ORB%), points scored the previous month, and percentage of wins this month. Participants' answers were collected using a multiple-choice question with 10 options. The choice options were composed of five different confidence level and two team options (Home Team or Away Team). Participants were forced to select only one option. An example of the options is displayed in Figure 3.14.

**Figure 3.14** Example of multiple-choice options in Experiment 8

- |   |  |  |
|---|--|--|
| It is a toss-up but I think that        | <input type="radio"/> Home team will win | <input type="radio"/> Away team will win |
| There is a better than even chance that | <input type="radio"/> Home team will win | <input type="radio"/> Away team will win |
| It is quite likely that                 | <input type="radio"/> Home team will win | <input type="radio"/> Away team will win |
| It is highly probable that              | <input type="radio"/> Home team will win | <input type="radio"/> Away team will win |
| I am almost certain that                | <input type="radio"/> Home team will win | <input type="radio"/> Away team will win |

**Procedure.** The task used in Experiment 8 was identical to Experiment 6 and 7. Participants predicted basketball game outcomes based on the four statistical measures. First, participants read about the task and definition of each basketball statistics, then completed a short training block comprised of three trials. The training blocks were identical to Experiment 6. In the first two training trials the two advisers predicted that the same team would win, and they were correct. The expert adviser gave the correct answer for all three training trials, but the novice's answer was incorrect in the last training trial. After each training trial, participants were provided with the winning team and score difference between the Home Team and Away Team. The game outcomes were

not provided in the test trials. Upon completing the training trials, participants engaged in test trials, which comprised of 48 trials split into two blocks. After the test trials, participants gave adviser competence rating and ranked each advisers' use of each basketball statistics from used the most to the least. Order of presentation of the two advisers for the rating and the ranking questions were randomised. Demographic information (age and gender) was then collected. Finally, participants were debriefed.

### ***Data Analysis***

The data analysis for Experiment 8 followed the same method as Experiment 6 and 7. Analysis was conducted separately for ambiguous and unambiguous evidence trials for adviser influence (change in participant confidence pre- and post-advice) and weight of advice (WOA) measure. Participants' linguistic expression of confidence had to be translated into a numerical value to calculate confidence distance (adviser confidence – participant's initial confidence) and WOA. Each linguistic expression represented a range of confidence. Therefore, following the same method used in Experiment 7, each expression was translated as the mid-point value of the range. For example, *Almost Certain*, which represents 90% to 100% confidence, was calculated as 95%.

### **3.4.2 Results**

***Participants' Accuracy.*** Similar to Experiment 6 and 7, in the ambiguous evidence trials, participants changed their initial answer more often following the expert's advice,  $t(59) = 4.46, p < .001, d = 0.58$ . The difference in initial and final accuracy was also greater for trials with the expert adviser than with the novice adviser,  $t(59) = -3.10, p = .003, d = 0.40$ .

**Table 3.4** Participants' accuracy by trial and adviser type in Experiment 8.

	Initial Accuracy (%)	Final Accuracy (%)	Agreeing Trials (%)	Changed Answer (%)
Ambiguous Evidence				
Expert	51	60	58	17
Novice	52	57	59	9
Unambiguous Evidence				
Expert	99	99	99	0
Novice	92	92	99	0

**Adviser Rating.** Consistent with Experiment 6 and 7, the expert adviser's competence rating ( $M_{exp} = 4.46$ ,  $SE_{exp} = 0.12$ ) was significantly greater than the novice's rating ( $M_{nov} = 2.68$ ,  $SE_{nov} = 0.13$ ),  $t(59) = 11.41$ ,  $p < .001$ ,  $d = 1.47$ , indicating that participants perceived the expert adviser to be more competent than the novice adviser.

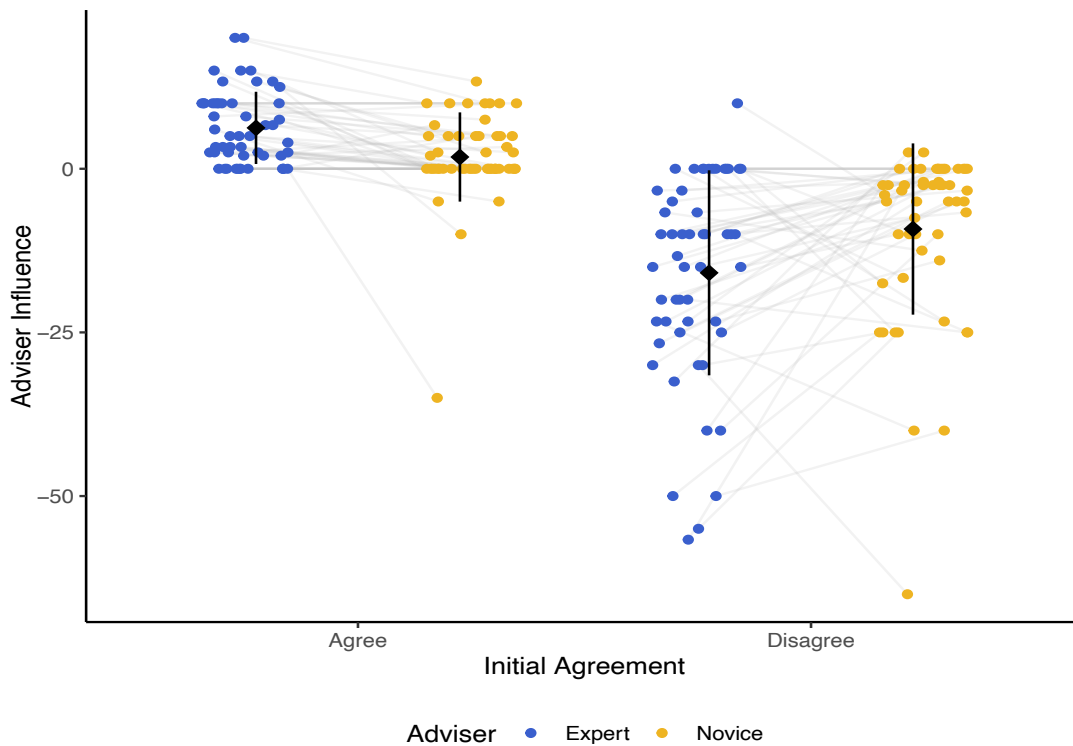
**Adviser Influence.** Participants' change in confidence from initial to final answer (i.e., adviser influence measure) was recorded to measure advice adoption for each adviser. The effect of adviser type on participants' confidence change was analysed separately for ambiguous and unambiguous evidence trials.

*Ambiguous evidence trials.* It was predicted that agreeing advice would increase confidence whereas disagreeing advice would lead to reduction in confidence, where this effect would be stronger following the expert's compared to the novice's advice (H1). To test H1, trials with ambiguous basketball statistics were analysed using two t-tests, ran separately for agreeing and disagreeing trials. It was predicted that adviser influence (change in confidence pre- to post-advice) would be positive when advisers agree with the participant's initial answer and negative under disagreement. The adviser influence was indeed greater than zero under agreement ( $M_{Agree} = 4.09$ ,  $SD_{Agree} = 5.30$ ),  $t(59) = 5.98$ ,

$p < .001$ ,  $d = 0.77$ . Whereas, adviser influence was below zero under disagreement ( $M_{Disagree} = -12.27$ ,  $SD_{Disagree} = 11.38$ ),  $t(59) = -8.35$ ,  $p < .001$ ,  $d = -1.08$

A 2 (agree vs. disagree) x 2 (expert vs. novice) repeated measure ANOVA was conducted with adviser influence as the dependent variable. Seven participants were excluded from the analysis due to incomplete data set in all four trial types. Therefore, data from 53 participants were analysed. Consistent with Experiment 7, there was a significant main effect of agreement,  $F(1,52) = 68.01$ ,  $p < .001$ ,  $\eta^2 = .36$ , but there was no significant main effect of adviser type,  $F(1,52) = 0.97$ ,  $p = .331$ ,  $\eta^2 = .003$ . Of critical interest, the interaction term was significant,  $F(1,52) = 17.55$ ,  $p < .001$ ,  $\eta^2 = .06$ . Pairwise comparison with Bonferroni correction showed that there was a significant difference in confidence change by adviser type in the agreeing ( $p < .001$ ,  $d = 0.72$ ) and disagreeing ( $p = .005$ ,  $d = -0.46$ ) trials. In the agreeing trials, participants increased their confidence more following the expert's advice ( $M_{exp} = 6.25$ ,  $SE_{exp} = 0.76$ ) compared to the novice's advice ( $M_{nov} = 1.80$ ,  $SE_{nov} = 0.94$ ). Similarly, when faced with disagreement with the advisers, participants decreased their confidence significantly more with the expert adviser ( $M_{exp} = -15.9$ ,  $SE_{exp} = 2.15$ ) compared to the novice adviser ( $M_{nov} = -9.20$ ,  $SE_{nov} = 1.80$ ).

**Figure 3.15** *Adviser influence following advice in ambiguous evidence trials in Experiment 8*

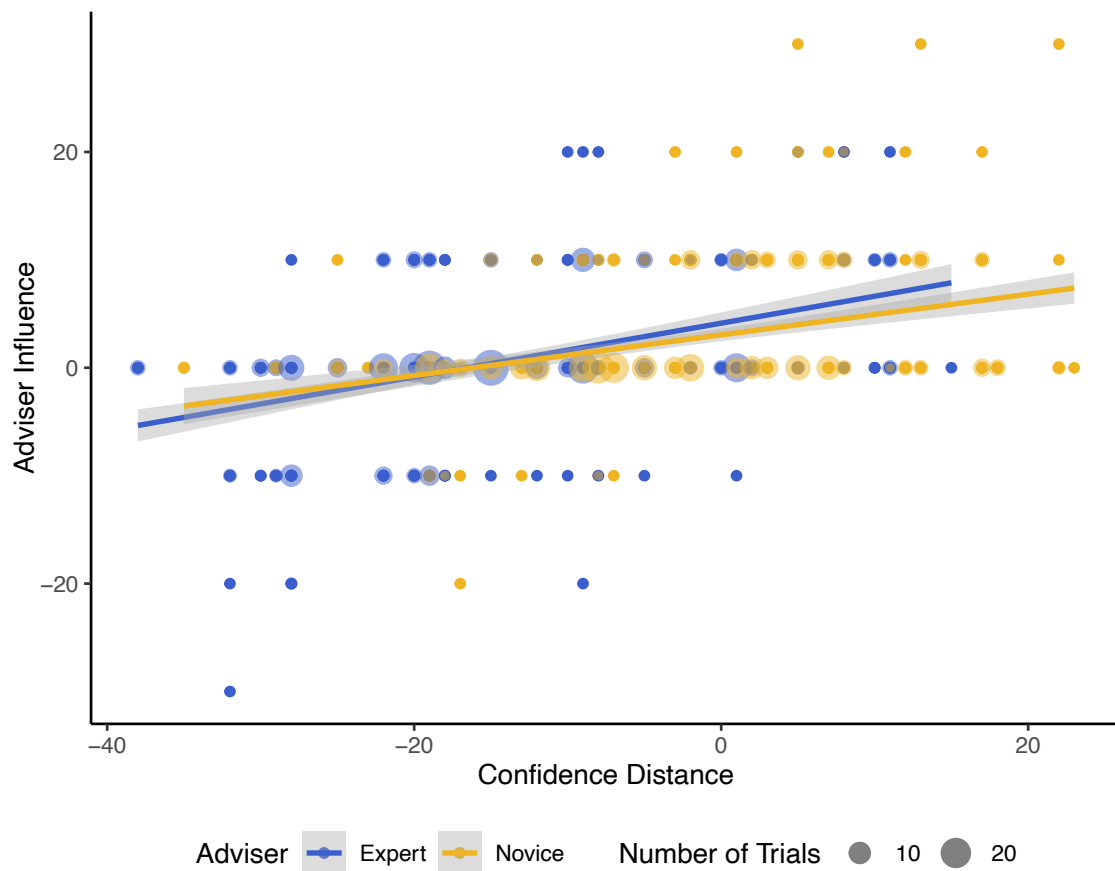


*Unambiguous evidence trials.* Unambiguous evidence trials were analysed to test H2, which hypothesised that agreement from uncertain advisers would lead participants to feel less confident, with this effect being greater for an expert than a novice adviser. There were only few trials where participants initial answer and advice disagreed. Therefore, only trials where participants initial answer and advice were in agreement were analysed. Data were analysed using t-tests. In contrast to Experiment 6 but consistent with Experiment 7, adviser influence ( $M = 1.52, SD = 3.51$ ) was significantly greater than zero,  $t(59) = 3.36, p = .001, d = 0.43$ . Moreover, the adviser influence value was significantly higher for the novice adviser ( $M_{Nov} = 2.40, SD_{Nov} = 3.71$ ) compared to the expert adviser ( $M_{Exp} = 0.67, SD_{Exp} = 4.14$ ),  $t(59) = -3.79, p < .001, d = -0.49$ .

In order to better understand the results, a linear mixed model was used, where adviser type and confidence distance (adviser's confidence - participant's initial confidence) were

the fixed effects. Participants were included as random effect. There was a significant main effect of confidence distance ( $b = 0.25, SE = 0.02, p < .001$ ). In line with Experiment 6 and 7, participants generally increased their confidence when confidence distance was positive and reduced their confidence when confidence distance was negative (despite advisers' agreement on which team would win). There was also a significant main effect of adviser type ( $b = 0.75, SE = 0.25, p = .002$ ). The interaction between confidence distance and adviser type was significant ( $b = 0.04, SE = 0.02, p = .025$ ), with the regression slope being significantly steeper for the expert adviser ( $b_{exp} = 0.29$ ) compared to the novice adviser ( $b_{nov} = 0.22$ ). This shows that participants were more sensitive to confidence distance with the expert adviser compared to the novice adviser and changed their confidence accordingly. Similar to Experiment 7, the overall positive adviser influence is likely because participants perceived advisers to have similar levels of confidence as themselves and thus became more confident when an adviser agreed. However, a notable difference with Experiment 6 and 7 is that there were more trials in the current experiment where participants did not change their confidence (i.e., adviser influence is 0). Additionally, in the present experiment, participants still decreased their confidence sometimes, but similar to Experiment 7, this occurred most often when the adviser was substantially less confident than the participant (Figure 3.16). This conservative change in confidence may be because participants' confidence was recorded using five linguistic expressions. Hence, participants may have only changed their confidence when they felt a considerable change in their confidence.

**Figure 3.16.** Adviser influence measure following advice in unambiguous evidence trials in Experiment 8

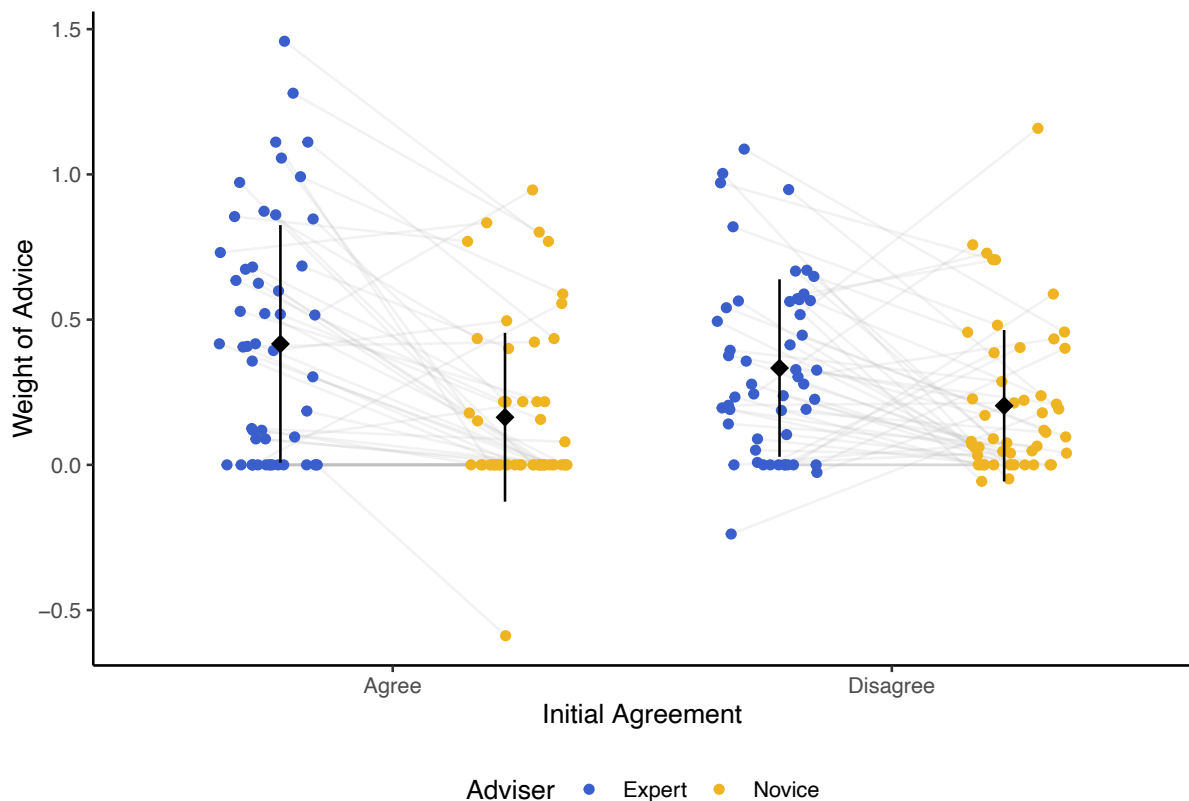


**Weight of Advice.** Weight of advice (WOA) was calculated as an additional measure of advice adoption. The participants and advisers expressed confidence using different methods (linguistic expressions vs. numerical values). Therefore, participants' linguistic expressions of confidence were translated to numerical values to calculate WOA. Hence, comparison of WOA values to values reported previous literature should be made cautiously.

**Ambiguous evidence trials.** A repeated measure ANOVA was carried out with agreement (agree vs. disagree) and adviser type (expert vs. novice) as the independent variables and weight of advice (WOA) as the dependent variable. Data from a total of 52 participants were analysed after eliminating participants with incomplete cases. The main effect of

agreement was not significant,  $F(1,51) = 0.22, p = .643, \eta^2 = .001$ . However, there was a significant main effect of adviser type,  $F(1,51) = 21.70, p < .001, \eta^2 = .08$ , where the expert adviser's WOA ( $M_{exp} = 0.38, SE_{exp} = 0.04$ ) was significantly greater than that of the novice adviser's WOA ( $M_{nov} = 0.18, SE_{nov} = 0.03$ ). The interaction between agreement and adviser type was also significant,  $F(1,51) = 6.67, p = .013, \eta^2 = .01$ . Pairwise comparison with Bonferroni correction showed that the difference between expert ( $M_{exp} = 0.42, SE_{exp} = 0.06$ ) and novice ( $M_{nov} = 0.16, SE_{nov} = 0.04$ ) was significant in agreeing trials ( $p < .001, d = 0.71$ ). This was also the case for disagreeing trials ( $M_{exp} = 0.33, SE_{exp} = 0.04; M_{nov} = 0.20, SE_{nov} = 0.04; p = .009, d = 0.46$ ). This suggests that the difference between expert and novice WOA is greater under agreement compared to disagreement (Figure 3.17).

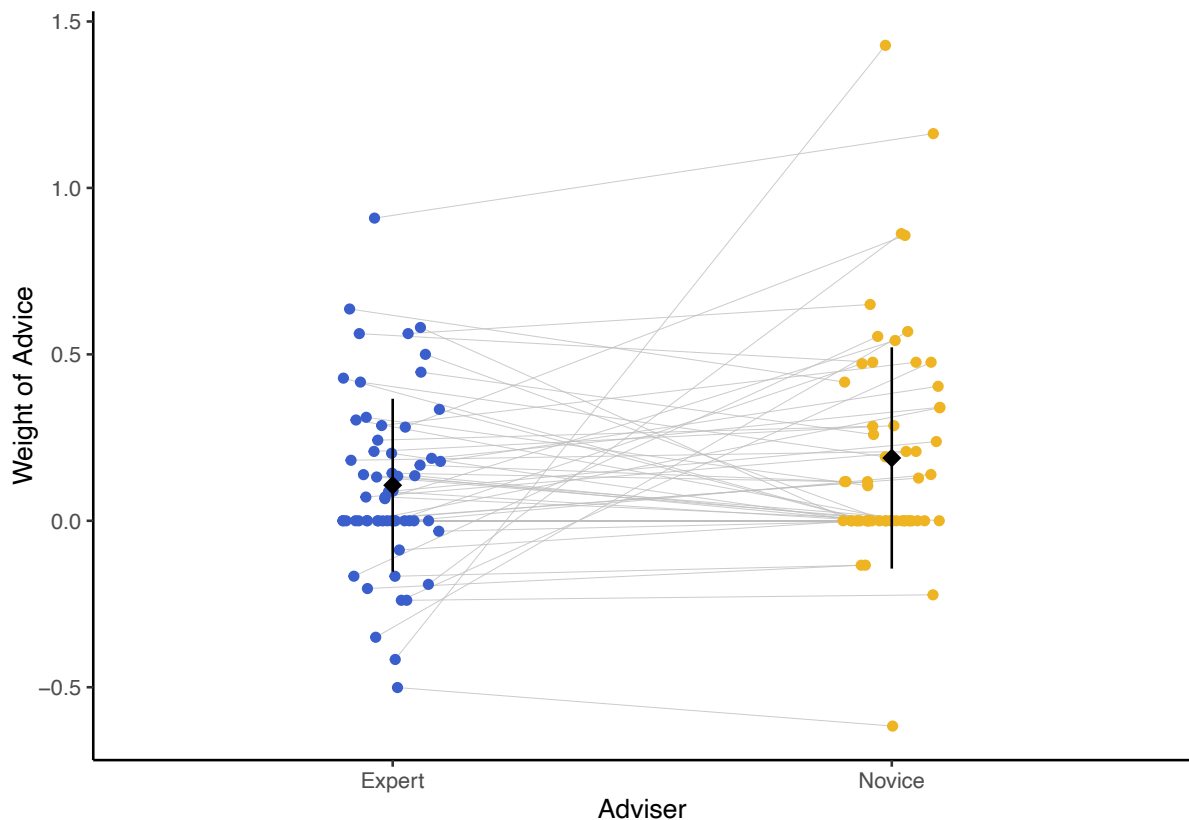
**Figure 3.17** *Weight of Advice of ambiguous evidence trials in Experiment 8*



*Unambiguous evidence trials.* A t-test was conducted to compare WOA after receiving the expert's versus the novice's advice. The WOA value was numerically higher for the novice adviser ( $M_{nov} = 0.19, SE_{nov} = 0.04$ ) compared to the expert adviser ( $M_{exp} = 0.11,$

$SE_{exp} = 0.03$ ) but the difference was not significant,  $t(59) = -1.67, p = .099, d = -0.22$ . Similar to Experiment 7, the difference between expert's and novice's WOA was driven by the greater number of negative WOA following the expert's advice (Figure 3.18). The negative WOA is likely a result of participants becoming more confident than the adviser following agreeing advice. In other words, despite negative confidence distance, participants sometimes increased their confidence (i.e., positive adviser influence), particularly after receiving the expert's advice (Figure 3.16).

**Figure 3.18** *Weight of Advice of unambiguous evidence trials in Experiment 8*

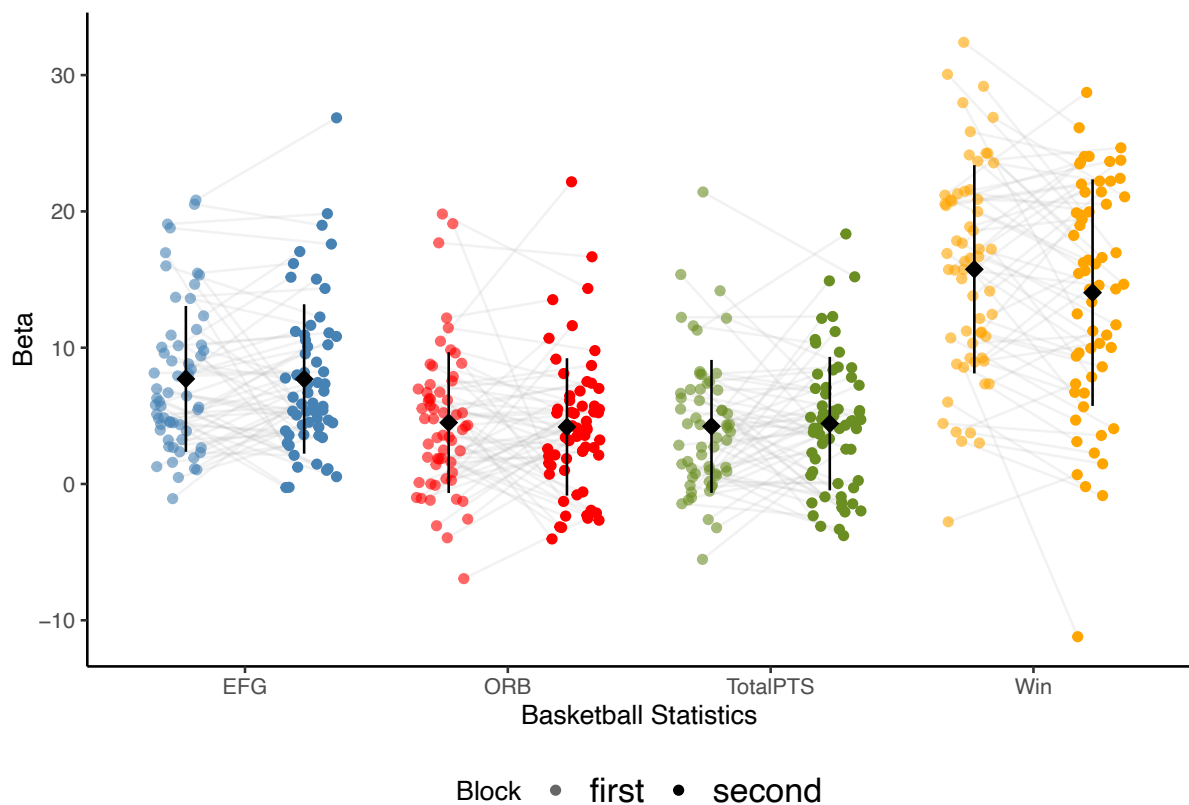


**Advisers' Variable Use Ranking.** In order to examine if participants were able to deduce the advisers weighting of the four basketball statistics, participants were asked to rank each adviser's use of the four basketball statistics from weighted the most to the least. Kendall's rank correlation coefficient was calculated between the participants' inferred ranking and the advisers' actual use of the basketball statistics, separately for the expert

and the novice advisers. The Kendall's ranking correlation coefficient between the participants inferred and expert adviser's actual ranking ( $M_{exp} = 0.22$ ,  $SE_{exp} = 0.07$ ) was significantly greater than the Kendall's rank correlation coefficient between participants' inferred and the novice's actual ranking ( $M_{nov} = -0.34$ ,  $SE_{nov} = 0.06$ ),  $t(59) = 6.68$ ,  $p < .001$ ,  $d = 0.86$ . This shows that participants were better at deducing the expert's ranking of the four basketball statistics compared to the novice's ranking.

**Participants variable use.** Hypothesis 3 predicted that unexpected adviser confidence would prompt participants to deduce and adopt adviser's weighting of the four basketball statistics. Therefore, participants' basketball statistics weighing was compared between the first and the second test blocks, following the same method as Experiment 6 and 7. This analysis assessed whether there are any changes in participants' use of the four basketball statistics in predicting the basketball game outcomes. Four separate t-tests were carried out (one for each basketball statistic) with the beta values calculated for each participant in each block. There was no significant difference in the use of effective field goal ( $p = .993$ ,  $d = 0.001$ ), offensive rebound ( $p = .655$ ,  $d = 0.06$ ), points scored the previous month ( $p = .767$ ,  $d = -0.04$ ). However, there was a significant decrease in the use of percentage of wins this month (weighted the least by the advisers) from the first block ( $M_{first} = 15.75$ ,  $SE_{first} = 0.99$ ) to the second block ( $M_{second} = 14.04$ ,  $SE_{second} = 1.07$ ),  $t(59) = 2.19$ ,  $p = .032$ ,  $d = 0.28$ . These findings partially replicate Experiment 6 and show some support for H3. Additionally, the trial effect was examined, which yielded the same results where only the beta value of the interaction between percentage of wins this month and trial order was significantly different from zero ( $M_{WIN} = -0.08$ ,  $SE_{WIN} = 0.04$ ),  $t(59) = -2.11$ ,  $p = .039$ ,  $d = 0.27$ .

**Figure 3.19** Change in use of four basketball statistics by block in Experiment 8



### 3.4.3 Discussion

In Experiment 8, the advisers gave their confidence using numerical values whereas participants were asked to use linguistic expressions to express their confidence. The experiment was conducted to confirm that participants' change in confidence was not a result of confidence matching.

When external information was ambiguous, the results were consistent with Experiment 6 and 7. Participants became more confident when advisers agreed with them. In contrast, when the advisers disagreed, participants became less confident in their prediction. The expert adviser had a stronger impact on people's confidence than the novice adviser. This was also confirmed in greater WOA of expert adviser compared to the novice adviser.

The results of unambiguous evidence trials in Experiment 8 only partially replicated Experiment 6 and 7. One difference compared to Experiment 6 and 7 was that participants

changed their confidence less. This was expected as there are only five levels of confidence. Therefore, participants are unlikely to change their confidence unless there is a large discrepancy between their own confidence and the advisers' confidence. Additionally, the overall change in confidence was positive. This contradicted results from Experiment 6 but was consistent with Experiment 7. However, in contrast to Experiment 7, the overall adviser influence was greater after the novice's advice than after the expert's advice. As visible in Figure 3.16, participants generally increased their confidence following agreement. However, they decreased their confidence when an adviser, and particularly an expert adviser, was considerably less confident than themselves. This would have led to the lower adviser influence value of the expert compared to the novice adviser.

In unambiguous evidence trials, the weight of advice (WOA) was higher for the novice compared to the expert adviser. This is unlikely to indicate that the novice adviser is more influential but driven by greater negative WOA values for the expert. The WOA value would be negative when an advisee becomes more confident than the adviser even though the advice expressed lower confidence than the advisee's initial confidence. It is possible that participants felt more confident in their initial answer following the expert's, rather than the novice's, advice. It is also important to note that, similar to Experiment 7, participants' linguistic expressions of confidence were translated to numeric values to calculate the WOA measure. However, the linguistic expressions represent a range of values and participants' actual confidence may differ from the assigned numeric value. Moreover, weight of advice is designed to measure advice adoption in estimation tasks. In this experiment, participants' confidence was collected rather than their estimation. Thus, participants' use of advice is likely to deviate from advice adoption in estimation

task. Additionally, as discussed in Chapter 1, WOA measure is unsuitable for multiple choice questions (MCQs). In the current experiment participants reported their confidence by selecting one of 10 options, which closely resembles MCQs. The current experiment demonstrates the limitations of WOA and thus, comparison with WOA values outlined in previous research should be made with caution.

Furthermore, it was hypothesised that participants would change their use of the four basketball statistics following repeated interaction with the advisers (H3). This was predicted as unexpected adviser confidence could lead participants to doubt their own weighting of the four basketball statistics. Therefore, the use of the four basketball statistics was compared between the first and second half of the trials. There was a significant decrease in the use of win percentage this month. This was the variable used the least by both advisers. Participants were not explicitly asked to guess the advisers' decision-making strategy. Nonetheless, it is possible that participants voluntarily engaged in deducing which variables the advisers prioritised and partially endorsed their strategy. The change in participants' strategy was greater in the current experiment compared to Experiment 7. This might be because the adviser used numerical values to express their confidence instead of linguistic expressions. The numerical values are more segmented and arguably more informative than five linguistic expressions. This would have facilitated deduction of advisers' statistics use and adoption of advisers' strategy. However, the change in participants weighting of the basketball statistics was less evident compared to Experiment 6. This might be because participants changed their confidence less which made it more difficult to capture participants' change in the use of the four basketball statistics.

### 3.5 Meta-analysis of effect size

The meta-analysis was conducted on the effect sizes obtained in each experiment (Experiments 6-8) using the *metafor* package (Viechtbauer, 2010) in R. The meta-analysis results are reported separately for ambiguous and unambiguous evidence trials.

*Ambiguous evidence trials.* Two t-tests were conducted, each t-test compared if adviser influence (difference in pre- and post-advice confidence) is significantly different from zero following agreeing or disagreeing trials.

The effect size of the two t-tests were compared across Experiments 6-8. Two random-effects meta-analyses were conducted, which showed that adviser influence was significantly different from zero following agreeing advice,  $d = 0.83$ ,  $SE = 0.09$ , 95% CI[0.64, 1.01],  $z = 8.74$ ,  $p < .001$ . This was also the case for adviser influence after disagreeing advice  $d = -1.15$ ,  $SE = 0.11$ , 95% CI[-1.35, -0.94],  $z = -10.91$ ,  $p < .001$ . There was no evidence of heterogeneity in either analysis (agree:  $Q(2) = 0.81$ ,  $p = .781$ ,  $I^2 = 0\%$ ; disagree:  $Q(2) = 0.47$ ,  $p = .792$ ,  $I^2 = 0\%$ ).

The effect size of the 2 (agree vs. disagree) x 2 (expert vs. novice) repeated measure ANOVA across Experiments 6-8 were also examined. The partial  $\eta^2$  values were first converted to Cohen's  $d$  for ease of comparison following Lakens (2013). A random-effects model was fitted using the *metafor* package (Viechtbauer, 2010) with restricted maximum likelihood (REML). The analysis found a significant effect of agreement,  $d = 1.64$ ,  $SE = 0.14$ , 95% CI[1.37, 1.91],  $z = 12.05$ ,  $p < .001$ . There was no evidence of heterogeneity,  $Q(2) = 0.94$ ,  $p = .625$ ,  $I^2 = 0\%$ . The random-effects model found a small, non-significant effect,  $d = 0.16$ ,  $SE = 0.09$ , 95% CI[-0.02, 0.34],  $z = 1.70$ ,  $p = .090$ . The

heterogeneity across experiments was not significant,  $Q(2) = 3.04, p = .219, I^2 = 6.97\%$ . The overall effect of the interaction between agreement and adviser type was large and significant,  $d = 0.58, SE = 0.10, 95\% CI[0.39, 0.77], z = 6.06, p < .001$ . No meaningful variability was found,  $Q(2) = 0.51, p = .777, I^2 = 0\%$ .

*Unambiguous evidence trials.* In the unambiguous evidence trials, whether overall adviser influence was significantly different from zero was tested using a t-test. A random-effects meta-analysis was conducted on the effect sizes (Cohen's  $d$ ) from Experiments 6-8. The analysis found that adviser influence was not significantly different from zero following agreeing advice,  $d = 0.29, SE = 0.21, 95\% CI[-0.12, 0.70], z = 1.37, p = .172$ . There was significant variability in the effect sizes,  $Q(2) = 9.94, p = .007, I^2 = 82.67\%$ . This was likely due to the negative mean adviser influence in Experiment 6 but positive adviser influence in Experiments 7 and 8.

The effect size of the linear mixed model was also compared. Cohen's  $d$  for each fixed effect (two main effects and one interaction term) was first calculated by dividing the unstandardised beta by the model residual SD. A random-effects meta-analysis showed a small but significant main effect of confidence distance,  $d = 0.05, SE = 0.004, 95\% CI[0.04, 0.05], z = 13.61, p < .001$ . There was moderate heterogeneity across experiments but this was not significant,  $Q(2) = 3.43, p = .180, I^2 = 44.57\%$ . The main effect of adviser type was also small but significant,  $d = 0.16, SE = 0.05, 95\% CI[0.07, 0.26], z = 3.42, p < .001$ . Once again, there was substantial heterogeneity but the Q-test was not significant,  $Q(2) = 5.29, p = .071, I^2 = 62.04\%$ . Of critical interest was the interaction between confidence distance and adviser type, which was significant but the pooled  $d$  was

negligible,  $d = 0.01$ ,  $SE = 0.003$ , 95% CI[0.003, 0.01],  $z = 2.97$ ,  $p = .003$ . There was no evidence of heterogeneity,  $Q(2) = 0.19$ ,  $p = .908$ ,  $I^2 = 0\%$ .

In addition, Experiments 6-8 also examined how participants' use of the four basketball statistics (effective field goal, offensive rebound, points scored the previous month, percentage of wins this month) changed from first half to the second half of the trials using four t-tests. Meta-analysis was conducted on the effect sizes (Cohen's  $d$ ) from each t-test. Random effects meta-analyses found no evidence to suggest that the use of eFG varied,  $d = -0.04$ ,  $SE = 0.08$ , 95% CI[-0.20, 0.12],  $z = -0.45$ ,  $p = .653$ . There was no evidence of heterogeneity,  $Q(2) = 0.76$ ,  $p = .685$ ,  $I^2 = 0\%$ . The pooled  $d$  of ORB was significantly different from zero,  $d = 0.17$ ,  $SE = 0.08$ , 95% CI[0.01, 0.33],  $z = 2.09$ ,  $p = .036$ . There was no evidence of heterogeneity,  $Q(2) = 1.30$ ,  $p = .522$ ,  $I^2 = 0\%$ . The effect size of both points scored the previous month ( $d = 0.16$ ,  $SE = 0.12$ , 95% CI[-0.08, 0.41],  $z = 1.31$ ,  $p = .191$ ) and percentage of wins this month ( $d = 0.22$ ,  $SE = 0.13$ , 95% CI[-0.03, 0.47],  $z = 1.75$ ,  $p = .081$ ) did not differ significantly from zero. For both points scored the previous month ( $Q(2) = 4.25$ ,  $p = .119$ ,  $I^2 = 53.18\%$ ) and percentage of wins this month ( $Q(2) = 4.35$ ,  $p = .113$ ,  $I^2 = 54.13\%$ ) there was moderate heterogeneity but the Q tests were not significant. This was unexpected as the use of percentage of wins this month significantly decreased in Experiments 6 and 8. The current results may be due to the variation in the effect sizes.

### **3.6 General Discussion**

The three experiments in this chapter examined how people process unexpected adviser confidence in relation to external information. Two types of advice-taking context were investigated. One was when the external information was ambiguous, but the adviser

expressed high confidence. In this context, advisees are likely to increase their confidence when an adviser agrees with their initial answer but lower their confidence when an adviser disagrees (H1). The main focus of the current experiment was to examine advisee behaviour in the second context where the external evidence was unambiguous, but the adviser displays low confidence. It was hypothesised that advisees would decrease their confidence despite an adviser agreeing with their initial answer (H2). Advice was given either by a novice or an expert adviser, where the expert adviser would have a greater influence on advisees (H1 & H2). Additionally, the unexpected adviser confidence might provoke advisees to reevaluate their assessment of external evidence and revise their decision-making strategy (H3). Three experiments were carried out, where the task was to predict the winning team of a basketball game given four pieces of basketball statistics.

In all three experiments, given ambiguous external information, advisees' confidence increased under agreement and decreased when advice disagreed with their initial answer, supporting existing literature (e.g., Julian et al., 1968; McGarty et al., 1993; Schultze et al., 2015). It was expected that people would significantly underweight the novice's advice as their high confidence would be taken as a sign of overconfidence and advisees tend to punish unwarranted high confidence (Rader et al., 2017; Sah et al., 2013; Tenney et al., 2008; Tenney et al., 2019; Van Zant, 2022; Vullioud et al., 2017). People did adhere less to the novice's advice compared to the expert's advice but still adopted the novice's advice to a considerable degree. For example, participants' confidence change in Experiment 6 was similar for expert and novice under agreement. A possible explanation is that people did not perceive the novice's high confidence as 'unwarranted'. A negative judgement of an adviser's high confidence must be accompanied by low accuracy (Tenney et al., 2008). However, since the evidence is ambiguous and actual accuracy

information is unavailable, people might have believed that the novice's high confidence stems from high accuracy. Alternatively, there are social aspects of advice-taking. For example, people believe that advice stems from one's intention to help (Harvey & Fischer, 1997) and there are expectations of reciprocity (Mahmoodi et al., 2018). Therefore, participants may have felt obliged to meet such social expectations by accepting advice. Furthermore, previous research has shown that advisees tend to somewhat accept advice (about 20% shift towards advice) even when advisers are inexperienced (Harvey & Fischer, 1997; Schultze et al., 2017). Schultze and colleagues (2017) argue that advice has an anchoring effect where advisees shift towards advice regardless of the perceived quality of advice.

The main advice-taking context of interest was when the external information was unambiguous, but the adviser displayed low confidence. It is generally expected that people increase their confidence when another person agrees with their opinion (Julian et al., 1968; McGarty et al., 1993; Schultze et al., 2015; Soll et al., 2022). However, the current chapter demonstrated that when an advisee is overconfident compared to the adviser, the advisee would often decrease their confidence even when the adviser agrees with their initial answer. In Experiment 6, participants overall reduced their confidence following agreeing advice, where the decrease in confidence following expert advice was significantly below zero. In Experiments 7 and 8, participants did generally become more confident after advice, but participants still decreased their confidence when an adviser was substantially less confident than themselves. This less frequent reduction in confidence compared to Experiment 6 might be due to the difference in how confidence was expressed. In Experiment 7, the adviser used linguistic expressions which represents a range of numeric values. Therefore, participants might have perceived adviser's

confidence as similar to their own confidence (i.e., advisees believed they were not overconfident). Thus, participants reduced their confidence less often and only when their overconfidence was apparent. In Experiment 8, participants reported their confidence using five linguistic expressions. Therefore, participants seldom changed their confidence and only did so when they felt a substantial change in confidence. Moreover, the decrease in confidence despite agreement in Experiment 7 and 8 is less likely to be driven by the possible desire to match one's confidence to the advisers' confidence (i.e., confidence matching) (Friedmann et al., 2024). Linguistic expressions represent a range of confidence and thus, participants are less likely to have believed that there was a precise "correct" confidence. Furthermore, as there was no exact numeric value to practice confidence matching, participants decrease in confidence are likely to reflect a genuine change in confidence.

Although this finding appears to deviate from previous literature, there are several potential explanations. Research suggests that evidence that is surprising has a greater impact on decision making as it grabs people's attention (Horstmann, 2015; Soltani & Izquierdo, 2019). The adviser's low confidence was likely unexpected for advisees given that the decision evidence was unambiguous, which might lead them to place greater weight on adviser's confidence and lower their confidence accordingly. The results could also be explained in terms of a more complex model of Bayesian updating. In a simple Bayesian model, agreement from an adviser would always lead to increase in confidence as it is evidence in favour of their initial judgement, and therefore increases probability of choice being correct. However, a more complex Bayesian model could suggest that people initially determine their Bayesian confidence according to the basketball statistics (i.e., input) (Bate, 2007; Meyniel et al., 2015). However, after receiving advice, people

may not simply perceive the adviser's confidence as another piece of evidence to adjust their posterior belief about the correct decision but could also update their beliefs about the task (e.g., difficulty of the task) (Kyburg 1987; Nassar et al., 2010). For example, the surprisingly low adviser confidence not only suggest that advisees mispredicted the game outcome but also that they have misjudged the complexity of the task and how to interpret the four basketball statistics. Hence, advisees might feel less confident in their general ability to perform the task (i.e., predict the winning basketball team). The results could also be considered in the metacognitive perspective. People may first learn about the basketball statistics and have a grasp of the definition of the information (simple cognition). An unexpected adviser confidence leads people to question their definition, and thus, evaluate their metacognition of the task. This leads to decrease in their perceived ability to evaluate the evidence and confidence subsequently (Fleming & Daw, 2017).

Additionally, how participants are using the four pieces of basketball statistics to determine the winning team was investigated. It was hypothesised that participants would use the advisers' unexpected confidences to determine the predictiveness of the four basketball statistics (H3). As discussed above, unexpected adviser confidence would decrease participants' confidence in their own decision and task performance. This would in turn lead participants to question their weighting of the four basketball statistics and encourage participants to learn the advisers' weighting of the basketball statistics. Results supporting H3 would also provide additional support for the other hypotheses (H1 & H2), as this demonstrates that participants are not simply confidence matching but using unexpected adviser confidence to inform their decisions. For example, consider the case where the external evidence is ambiguous, but the adviser is highly confident that Home Team would win. In this case, the advisee can examine which basketball statistics favour

Home Team and can learn that the adviser places higher value in these basketball statistics. Advisers' low confidence when external evidence is unambiguous also provides clues regarding the adviser's decision-making strategy. The adviser would be less confident when indicative basketball statistics only narrowly differs between the Home and Away Team. Therefore, advisees could compare the magnitude of difference for each basketball statistics and deduce which basketball statistics are assigned greater weight by the adviser. In the current chapter, some evidence was found that people modified their decision-making strategy over time to resemble the advisers' strategy. Specifically, their weighting of the variables used less by the advisers decreased. The decrease in use of these variables was significant in Experiment 6 and 8 and numerically lower in Experiment 7. However, this was not supported in the meta-analysis which found that the overall effect size was not significantly different from zero. This was likely due to the high variability in the effect sizes. On the other hand, participants did not increase the use of variables prioritised by the advisers. In fact, the meta-analysis of effect-sizes found that participants decreased the use of ORB, which was a variable prioritised by the advisers. This limited adoption of advisers' strategy may stem from two reasons. Firstly, advisees might have been unable to deduce the advisers' ranking of the four basketball statistics. Past research suggest that people have difficulties interpreting quantitative data and inferring quantitative relationships between cues (Dawes & Corrigan, 1974; Dawes et al., 1989). This was also evident in participants' limited ability to predict the advisers' weighting of the four basketball statistics. Another possible explanation is that participants might have been reluctant to fully adopt the advisers' strategy. Past research generally find that people place greater weight on their own opinion (Bailey et al., 2023; Soll & Larrick, 2009; Yaniv & Kleinberg, 2000). Thus, participants might have decided generally to maintain their own strategy.

However, the results still demonstrate some learning, which is surprising as participants were not prompted to learn the advisers' decision-making strategy nor were they provided with feedback on accuracy. There are indeed pragmatic benefits of adopting an adviser's strategy (i.e., adopt prior) especially when the advisee has limited knowledge (Kyburg, 1987), which was the case in the current experiments (e.g., most participants reported having limited knowledge in basketball in Experiment 6). Additionally, researchers have found that providing probability information helps people better learn cue relationships (Little & Lewandowsky, 2009). Adviser confidence might have played a similar role in the current experiment and was considered as a type of feedback. Participants could have perceived the proximity of their confidence to the advisers' confidence as how accurately they have assessed the predictiveness of the four basketball statistics. Furthermore, feedback is not always necessary for learning to occur as people are capable of learning by using their metacognition (Veenman et al., 2006). Adviser confidence is qualitatively different from objective feedback thus the cue learning effect is weak, but the effect was still present. This is an interesting area for future research, where the impact of varying feedback reliability on degree of cue learning could be investigated.

### **3.5.1 Limitations**

In the experiments, the communication was one sided where participants only received advice and did not have a chance to interact with the adviser. Nor was the adviser aware of how much their advice was used. However, in real-life there is greater communication and intricate social aspects involved in advice-taking (Rader et al., 2017). For example, in the experiments, there were no stakes in rejecting advice whereas, in reality, advisers perceive advisees who reject their advice as less competent (Palmeira & Romero Lopez, 2023), which reduces adviser's future advice-giving intention (Belkin & Kong, 2018;

Blunden et al., 2019) and could sour the relationship (Rader et al., 2017). The current methodology was used as the purpose of the experiments was to examine how much advice people privately and willingly adopt but advice adoption might be greater in real-life advice-taking contexts due to the social aspects at play. Similarly, advice was given in written form in the studies. There is no concern for this reducing advice adoption as written advice leads to high levels of advice adoption (Van Swol & Snizek, 2005). Yet, written advice still qualitatively differs from face-to-face interaction which may contain greater information through non-verbal cues (Walther et al., 2005). This may help advisees better detect the predictive power of each of the basketball statistics. Lastly, in the studies, how to evaluate the evidence provided (i.e., basketball statistics) was apparent. Participants were clearly told that higher numbers are desirable. However, external information is often ambiguous in real-life and could be difficult to evaluate (as good or bad) and how the piece of information affects judgement could be unclear until post-judgement. Therefore, it would be beneficial to conduct future research incorporating the complexity of real-life advice-taking.

### **3.5.2 Conclusions**

The current research shows that adviser confidence has a unique impact on decision-making. People do not simply treat advice as another piece of information but process it extensively. For instance, adviser confidence is used to make general assessments regarding the advice-taking context. In the current experiments, participants sometimes reduced their confidence despite agreement from the adviser, suggesting that people are using adviser confidence in two ways. First, to evaluate their immediate judgement such as if they guessed the winning basketball team correctly with appropriate level of confidence. The second is to assess if their evidence evaluation is correct and update their

prior assessment of the external information (i.e., basketball statistics). Hints of this behaviour were observed as people's weighting of the four basketball statistics shifted to somewhat mirror the advisers' decision-making strategy.

Previous literature has suggested that people process adviser confidence in relation to a variety of factors. Yaniv and Kleinberger (2000) has suggested that one of the reasons for egocentric discounting (placing greater weight on one's own opinion compared to advice) is due to the inaccessibility to the adviser's reasoning. This in turn indicates that increased access to the adviser's decision-making process would increase advice adoption (Van Swol, 2011). This suggests that people are not simply taking advice at face value but are trying to deduce or at least are curious about the adviser's thoughts. Additionally, people seem to use adviser confidence to make other behavioural or dispositional judgements about the adviser. Vullioud and colleagues (2017) argue that unwarranted over-confidence is punished harshly (compared to a competent adviser who is wrong) because advisees perceive high confidence as a sign of strong commitment to the advice. Furthermore, advisees stop trusting an adviser after discovering that their high confidence is miscalibrated with accuracy (Sah et al., 2013; Tenney et al., 2008; Tenney et al., 2019; Van Zant, 2022). This mistrust even extends to domains that is unrelated to the original advice-taking context (Vullioud et al., 2017). This suggests that people assume that an adviser has other negative traits when they are overly confident. In addition, people prefer advisers to express their confidence in relation to the advice-taking context. For instance, advisers who express some uncertainties are preferred when there is also uncertainty in the advice-taking context and the correct outcome is unclear (Gaertig & Simmons 2018; Gaertig & Simmons, 2023). Moreover, Soll & Larrick (2009) also propose that when given advice, people either abandon their own opinion and fully adopt advice or use an

averaging strategy. Which approach is taken is affected by how close the advice is to the advisee's initial opinion (Soll & Larrick 2009; Soll & Mannes, 2011; Van Swol, 2011). This shows that people evaluate advice in relation to one's own opinion. The current research adds to this line of work by exhibiting that people use adviser confidence to evaluate the adviser and the advice-taking context.

# Chapter 4: Learning External Information Validity from Adviser Confidence

---

## 4.1 Introduction

The main objective of the experiments in Chapter 3 was to examine the effect of unexpected adviser confidence on advisee's answers and their accompanying confidence. However, these experiments also provided preliminary evidence of overall change in participants' decision-making strategy following advice. More specifically, participants exhibited signs of learning the validity of external information from advisers' answer and confidence. Therefore, the experiments in the current chapter investigated this effect of learning from adviser confidence further.

There are different methods to evaluate if one's decision-making strategy is optimal. One method is to learn retrospectively based on the outcome of the decision. For example, in reinforcement learning, feedback on accuracy is provided which helps people learn about the environment (Sutton & Barto, 1998). Similar results are found in the cue learning literature. In cue learning tasks, participants are generally given multiple cues and are asked to guess the outcome where performance feedback is provided after each decision. After repeated trials, people can learn the relationship between cues and outcomes (Droll et al., 2009; Lagnado et al., 2006).

However, feedback on accuracy may be ambiguous or not always available in real-life. Nevertheless, feedback may not always be necessary for learning to occur. For example, researchers have found that reinforcement learning can still occur in the absence of

feedback (McKee & Westhe, 1978; Karni & Sagi, 1991). The importance of feedback is less clear in multiple cue learning. Some researchers find that providing full feedback (i.e., feedback on all decisions) results in the best learning (Droll et al., 2009). However, studies that have provided sparse or conditional feedback (e.g., feedback only when correct) found that people are still able to learn cue validity (De Klerk, 1972; Hendrix & Dudycha, 2015). Additionally, Stewart and colleagues (2011) suggest that full feedback leads to the fastest learning but the difference with incomplete feedback attenuates after repeated trials. These findings demonstrate that feedback on accuracy is not essential to learn cue relationships.

One thing people could use instead of accuracy feedback is confidence of others. If someone displays strong confidence about their decision, others will assume that their decision would yield a positive outcome. This is a plausible assumption as confidence represents the expectation of the extent to which a decision, opinion, or recommendation is correct (Klayman et al., 1999). Furthermore, confidence could be informative due to its link to various factors and behaviours. For example, people use their own confidence to assess the competence of others (Pescetelli & Yeung, 2021). Confidence also relates to one's decision to seek further information (Desender et al., 2018; Locander & Hermann, 1979) and advice (Pescetelli et al., 2021). Furthermore, people use confidence to guide their behaviour such as which task to complete first (Aguilar-Lleyda et al., 2020). Additionally, people make diverse and complex inferences about adviser confidence. For example, people assume that high levels of confidence also indicate greater competence (i.e., *confidence heuristic*) (Price & Stone, 2004). Advisees also tend to view an adviser's confidence as their degree of commitment to their answer (Vullioud et al., 2017). Moreover, confidence is strongly related to external information. One's confidence is not

static, but changes dynamically based on the accumulated information (Desender et al., 2021). Hence, people are likely to also evaluate adviser confidence in relation to the available information. For example, people perceive an adviser who deviates from external cues of quality (e.g., price) to have lower accuracy but greater expertise (Palmeira, 2020).

These findings suggest that people not only use adviser confidence to assess the accuracy of advice but also make various inferences about the external environment. Therefore, advisees may also use adviser's confidence to assess the accuracy of their own decision. This would facilitate learning of external information validity, similar to accuracy feedback in reinforcement learning and cue learning. Furthermore, researchers have found that providing probabilistic feedback with lower accuracy in lieu of unvarying accuracy feedback led to better learning of the overall relationship between cues (Little & Lewandowsky, 2009). Therefore, adviser confidence might be a more helpful tool to learn external information that is intertwined, compared to simple accuracy feedback. Indeed, experiments in Chapter 3 showed signs that people were changing their weighting of the external information to better resemble the adviser's weighting after repeated advice-taking. This might be because people were able to learn the complex relationship between the basketball statistics and their validity from adviser confidence.

In Chapter 3, this learning from adviser confidence might have occurred as the adviser confidence was informative. For adviser confidence to be useful for learning the external information, there must be some discrepancy between the external information and adviser confidence, similar to the test trials in Experiment 6-8. It would be difficult to assess the relationship between the external information and the outcome, if all external

information suggests the same outcome (e.g., all basketball statistics high for Home Team). For example, imagine that two basketball statistics are higher for Home Team and the remaining two are higher for the Away Team. If an adviser predicts that Home Team would win, advisees can infer that the adviser prioritises the first two pieces of information in their decision-making. Through similar repeated interaction, advisees would be able to deduce the advisers use of the external information and decision-making strategy. In Chapter 3, participants showed signs of undergoing this process during advice-taking. However, this was not the main objective of Chapter 3 and was not formally investigated. Therefore, in the current chapter, learning of external information validity from adviser confidence will be investigated further. Hence, the following hypothesis will be tested.

H1: Advisees will be able to accurately assess the adviser's use of the external information in predicting outcomes.

Researchers have found that people often learn and adopt others' behaviour. This tendency increases when people are unsure about their own decision (Lee & Dry, 2006; Morgan et al., 2015; Toelch et al., 2014). In uncertain situations, people disregard the information they have acquired and adopt social information, albeit cautiously (Toelch et al., 2014). Additionally, people are more likely to comply with social information when the external environment is complex where the decision-outcome is dependent on various intertwined factors (Rendell et al., 2010). This is conceivable as presence of a large number of potentially predictive factors would cloud the relationship between external information and outcome, which increases one's decision uncertainty and subsequent bias towards adopting social information.

In fact, it is logical to adopt behaviours of others considering the various benefits. For instance, it reduces the time spent deciding and the risks that arises from trying different options (i.e., trial-and-error) (Izquierdo et al., 2018). Indeed, researchers have found that various animals' likelihood to imitate another individual increased when there were high costs of obtaining the information themselves (Valone, 2007). Additionally, learning the strategies used by better performing others ensures improved decision-making (Izquierdo et al., 2018). Research on social learning suggest that each individual's independent decision-making may be beneficial for the group but exploiting the strategies used by others lead to greater accuracy at an individual level (Lazer & Friedman, 2007; March, 1991; Mason & Watts, 2012). Furthermore, it may even result in greater decision accuracy for groups to follow one individual when the individual has greater expertise compared to the rest of the group members. More specifically, when a group is performing an estimation task (e.g., guessing height of a mountain), converging towards the best performing individual leads to higher decision accuracy rather than taking the average of everyone's answer (King et al., 2012).

Adhering to social information is also more beneficial when there are numerous factors associated with the decision. For example, a study pitted different decision-making strategies in a multi-armed-bandit task with 100 arms with varying payoffs. Surprisingly, the strategy that relied on adopting others' strategies was the most successful (Rendell et al., 2010). In this context, efforts to deduce the reward of each arm would be time consuming and costly. Instead, by observing others, it is possible to weed out the bad strategies and find the optimal strategy that maximises reward. In the real-world, there is an abundance of information which can be useful but also overwhelming to explore. The complexity and uncertainty of the environment would lead people to rely on other's

decision-making strategy. In particular, learning from an expert would help people make better decisions and increase their accuracy than persisting with their own strategy. Therefore, the following hypotheses are proposed:

H2: Advisees will adopt an adviser's use of the cues after repeated interaction.

H3: Adoption of the adviser's use of the external information will lead to increased accuracy in decision-making.

#### **4.1.1 Research Outline**

The task from Chapter 3 was also used in the current chapter, where participants were given four basketball statistics and predicted the winning team. Identical to Experiment 6-8, the validity of the four basketball statistics varied (refer to Experiment 6 for further details). The objective of this chapter was to investigate if participants could learn which statistics are the most predictive from adviser confidence and modify their weighting of the statistics accordingly. Although this was not the main objective, Experiments 6-8 found preliminary evidence of learning from adviser confidence, where participants reduced weighting of less predictive cues (that were accordingly down weighted by an expert adviser). Therefore, the current experiments' aim was to formally examine this learning process, with some methodological changes intended to enhance learning effects. More specifically, there was only one (expert) adviser, and the experiments were longer (96 trials instead of 48 trials). Additionally, there was a discrepancy between external information ambiguity and adviser confidence in all trials (e.g., external information ambiguous but adviser is highly confident or external evidence unambiguous, but adviser is uncertain). It was expected that these changes would make all trials informative and thus, greater opportunity to learn the adviser's weighting of the basketball statistics. In

Experiment 9, all trials followed the judge-adviser-system (JAS), where participants first guessed the winning basketball team alone and then received advice. After, participants could update their answer. To preview the results, comparing the first and second half of the trials revealed no signs of change in participants' weighting of the four basketball statistics in Experiment 9. Therefore, in Experiment 10, the methodology was modified to better resemble methods used in cue learning literature. In the first and last 24 trials, participants predicted the winning team without any advice. In contrast, the middle 48 trials followed JAS. Additionally, the benefits of learning from the adviser were highlighted by providing the adviser's and participant's accuracy and there were incentives for accuracy. Despite the methodological changes, no signs of learning were observed in Experiment 10 either.

## **4.2 Experiment 9**

The main goal of the experiment was to investigate if people are able to infer the validity of external information from adviser confidence. Literature in cue learning and my previous experiments (Experiments 6 – 8) suggest that learning can still occur from adviser confidence in the absence of objective feedback. Experiment 9 investigated whether repeated interaction with an adviser helps participants learn about the external information and their relevance in decision-making. More specifically, using the task from Chapter 3, participants viewed four basketball statistics and predicted the winning team. Participants first submitted their initial answer and then received advice. It was predicted that the advice provided will act as a proxy for accuracy feedback. This would help participants determine the informativeness of each basketball statistics. Participants weighting of the four basketball statistics in their initial decision was compared between the first and second half of the trials.

### 4.2.1 Methods

**Participants.** A total of 30 (15 = Female, 15 = Male,  $M_{age} = 34.7$ ,  $SD_{age} = 7.2$ ) participants were recruited through Prolific (<https://www.prolific.com>). Before commencing the experiment, all participants gave informed consent in accordance with the University of Oxford Central University Research Ethics Committee guidelines. Participants were compensated £9/hour based on the median duration of the experiment.

**Task.** The task used in this experiment was similar to the task in Experiment 6-8 with small changes. Participants were asked to predict which team would win a basketball match (i.e., binary decision-making), once again based on four statistics provided for each team: Effective Field Goal (eFG%), Offensive Rebound (ORB%), Total Points Scored the Previous Month and Win Percentage This Month. The basketball statistics were explained in laymen's terms, which was identical to the explanations given in Experiment 6-8 (refer to Figure 3.1). Therefore, participants were not required to have prior knowledge in Basketball.

The task followed the judge-adviser system (JAS) (Sniezek & Buckley, 1995). Participants first viewed the four basketball statistics alone and gave their initial answer. They were asked to indicate which team would win and their accompanying confidence level on a horizontal continuous scale ranging from *100% Home Team* to *100% Away Team*. The slider starting point was the mid-point (50%). Participants were required to move the slider and give their answer to proceed. After submitting their initial answer, an adviser appeared and gave their answer and confidence level. In Experiments 6-8, participants either viewed an expert or a novice adviser, but here only the expert adviser appeared and was called 'James' instead of 'Expert'. While viewing the advice,

participants were asked to indicate which team would win and by how many points. This was another major difference from Experiments 6-8, which asked for participants' confidence post-advice. Answers were once again collected on a horizontal scale, but the scale ranged from *Home team wins by 50 points* to *Away team wins by 50 points*. The slider starting point was once again the mid-point of the scale (0 points). An example of a trial is shown in Figure 4.1.

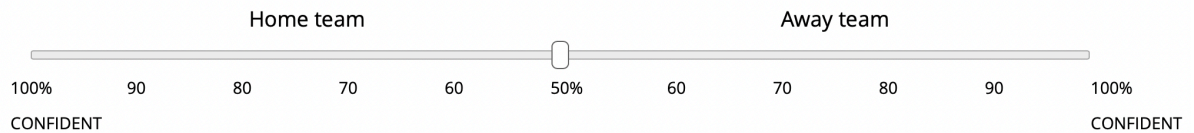
This change to requesting score prediction post-advice was done in anticipation of running another experiment in which participants would receive objective information (i.e., the winning team) about the game result (rather than advice) after making their initial prediction. In this case, it would be illogical to request their predicted winning team (as winning team was known) and confidence again. Instead, predicted score difference was asked post-advice/feedback, which could be requested and compared across experiments using advice and objective feedback. This experiment was planned in order to compare learning following advice versus objective information. However, comparison between advice and feedback was not carried out because, as previously mentioned, no signs of learning from advice were observed.

**Figure 4.1** Example trial of Experiment 9 (pre- and post-advice)

Please read the information about each team carefully and indicate which team you think will win using the slider.

Team statistics	Home team	Away team
Shooting accuracy (eFG%)	58.7	49.4
Offensive rebound (ORB%)	8.8	39
Points scored last month	110.4	105.8
Percentage of wins this month	27.3	45.5

Which team do you think will win? Home team or Away team?



You must move the slider and answer the question to continue.

Continue

Please read the information about each team carefully and indicate how many more points the winning team scored using the slider.

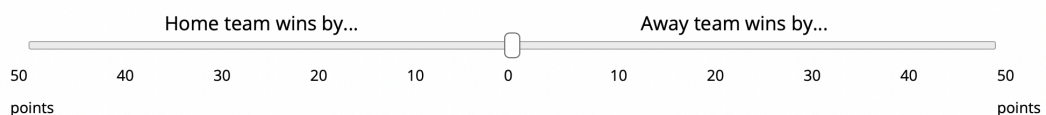
Team statistics	Home team	Away team
Shooting accuracy (eFG%)	58.7	49.4
Offensive rebound (ORB%)	8.8	39
Points scored last month	110.4	105.8
Percentage of wins this month	27.3	45.5

I am 56% confident that Home team will win.



James

What is your prediction of the final score?



You must move the slider and answer the question to continue.

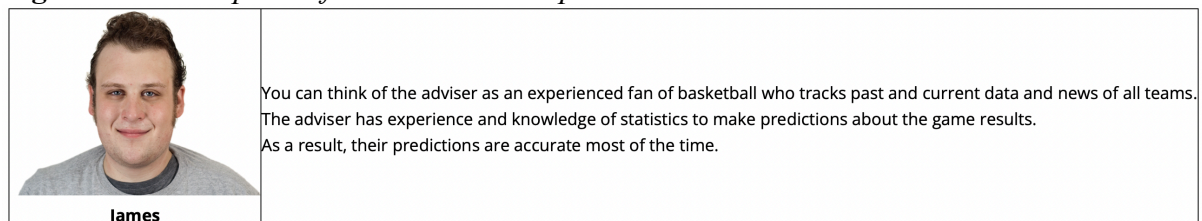
Continue

The purpose of the experiment was to test if participants can learn the validity of the external information from adviser confidence. Therefore, basketball games that were informative were selected as stimuli for the trials, which resembles the test trials in Experiments 6-8. More specifically, there was a discrepancy between the external information ambiguity and adviser confidence. In some trials, the adviser displayed high confidence despite ambiguous external information (e.g., only two out of four basketball

statistics are higher for Home Team). In other trials, the external information was unambiguous (e.g., all measures higher for Home Team), but the adviser expressed uncertainty. Out of the originally collected 168 basketball game data, 96 games were newly selected as stimuli for Experiment 9. One key difference with Experiments 6-8 is that there were no filler trials where the adviser confidence matched the external information ambiguity (e.g., adviser displays low confidence when only two statistics are high for Home Team).

**Adviser.** In standard cue learning experiments participants are often given feedback on accuracy (Lagnado et al., 2006). In the current experiment, this was replaced with advice and thus, it was essential that participants heeded to advice. Therefore, the adviser was described to have high levels of expertise to increase trust and reliance on advice. This description was identical to the description of an ‘Expert’ adviser in Experiments 6-8. Additionally, to make the adviser more realistic, the adviser was represented with a real-life picture (used in Experiments 6-8) and was given a name (James). The picture was of a white male smiling with his mouth closed, which was taken from the Chicago Face Database (<https://www.chicagofaces.org>). The picture of the adviser and the description are shown in Figure 4.2.

**Figure 4.2** *Description of the Adviser in Experiment 9 and 10*



The adviser’s answer and confidence were meaningfully calibrated based on the informativeness of the basketball statistics following the same methods as Chapter 3.

More specifically, a linear regression was carried out with the basketball statistics as the predictors and the game outcome (i.e., difference in score) as the dependent variable. The adviser's predictions of the score were created from the obtained regression coefficients. This was then translated into the adviser's confidence, where greater score difference led to greater confidence. A detailed account of the linear regression process is outlined in the Methods section of Experiment 6.

***Procedure.*** Participants first received instructions about the task and explanation about the four basketball statistics, then were given a practice question. Participants were required to move the slider and give an answer to proceed. After, an adviser gave their predicted winning team and confidence level. The slider appeared two seconds later than the rest of the information on the page to ensure that participants read the advice before answering. Once participants submitted their final answer, the game outcome was given, outlining the winning team and score difference.

Next, participants completed a training block of 3 trials. In each trial, participants first submitted their initial answer then received advice. Once again, the slider appeared two seconds later. Upon submitting their final answer, the game outcome (i.e., winning team and score difference) was given. This was to help participants gauge their own performance and highlight the accuracy of the adviser. Prior to commencing the test trials, participants were once again given the explanation of the basketball statistics to help refresh their memory and informed that the game outcomes will no longer be provided.

There were four test blocks where each block had 24 trials (i.e., 96 trials in total). In all trials, participants first viewed the team statistics alone and gave their predicted winning

team and confidence level using the slider. After submitting their initial answer, advice was provided on the next page where the slider appeared two seconds later. Participants then gave their final predicted winning team and score difference. Within each block there was one attention checking question, where participants were asked to move the slider to 100% Home Team (far left). Participants were allowed a 10-point error and anything above 90% Home Team was recognised as correct. The experiment terminated for participants who failed this question. Participants were allowed to take breaks in between blocks and proceed at their own pace. During the block breaks, the explanation of the basketball statistics was once again provided as a reminder.

Once participants completed all test blocks, they were asked to estimate which basketball statistics the adviser prioritised. Participants were asked to rank the statistics from used the most (= 1) to the least (= 4). Upon submission, participants were also asked to give a ranking of their own use of the basketball statistics. Participants' ranking was collected following the same format as the adviser's ranking. Next, the perceived competence of the adviser was collected using the same questionnaire as Experiments 6-8, which was extracted from Gaertig and Simmons (2018). The questionnaire comprised of seven questions and answers were collected on a 7-point Likert-scale (1 = *Not at all*, 7 = *Extremely*). The score for each question was averaged to create a final adviser competence score. Finally, participants gave their demographic information (age, gender). All participants were debriefed at the end of the experiment.

### ***Data Analysis***

The objective of Experiment 9 was to investigate if advisees can learn the validity of the external information from adviser confidence and modify their own weighting.

Participants' learning was examined using three different approaches. First, whether participants could deduce an adviser's use of the external information was investigated (H1). This was tested through participants' estimated ranking of the adviser's use of the four basketball statistics. Participants' estimated ranking will be compared to the adviser's actual ranking by calculating a Kendall's rank correlation coefficient. The average correlation coefficient across participants will be compared to zero using a t-test to examine if participants' estimation is superior to random guessing. In addition to the stated hypotheses, participants' own ranking of the four basketball statistics and participants' estimation of the adviser's ranking will be compared. More specifically, another Kendall's rank correlation coefficient will be calculated and tested against zero through a t-test. This was to examine if there are any overlap between the two ranking.

The second approach to testing learning was to investigate if participants' use of the four basketball statistics changes following repeated interaction with an adviser (H2). Therefore, the use of the four basketball statistics in the first half (i.e., first 48 trials) and the second half (i.e., final 48 trials) were compared. For each set of 48 trials, a linear regression was conducted with participants' initial answer as the dependent variable. Participants' answers were translated to a continuous scale ranging from 0 (= 100% home team) to 100 (=100% away team). The four basketball statistics were included as the predictors in the linear regression, where the linear regression was conducted separately per participant. Thus, two sets of beta values were derived for each participant – one set for the first half of the trials and another set for the second half of the trials. These beta values were compared through four paired t-tests (one for each basketball statistics). The linear regression focused on participants initial answer rather than post-advice answer. This was because the main interest of the experiment was to examine how participants

changed their own decision-making strategy (i.e., weighting of the four basketball statistics). Participants' change in weighting would suggest learning from adviser confidence as no performance feedback was provided.

The third test of learning was to compare participants' initial answer accuracy in the first and second half of the trials (H3). This was to inspect if participants became more accurate as they learned about the adviser's weighting of the four basketball statistics and shifted their own weighting. Brier score was calculated following the equation outlined below, where  $f_t$  indicates the forecasted probability (ranges from 0 to 1) and  $o_t$  represents the outcome (either 0 or 1). Brier score values range from 0 to 1, where 0 indicates perfect prediction of the outcome whereas 1 suggests wrongfully made prediction with full confidence. Brier score was used to measure participants' accuracy instead of tallying correctly guessed winning team. This was because Brier score takes into account the accuracy of the probability (i.e., confidence) estimation. However, participants' accuracy in binary game outcome prediction will also be reported for comparison.

$$Brier\ Score = \frac{1}{N} \sum_{t=1}^N (f_t - o_t)^2$$

The original plan of the experiment was to compare change in participants' use of the basketball statistics following objective feedback, expert's advice and novice's advice. Therefore, adviser rating was collected to ensure that the competency rating was higher for the expert adviser (current experiment). However, because no change in the use of the four basketball statistics was observed in this experiment, the parallel experiment with a novice adviser was not conducted. Therefore, adviser rating ( $M = 4.06$ ,  $SD = 1.12$ ) was not analysed.

#### 4.2.2 Results

**Ranking Correlation.** Participants were asked to estimate which basketball statistics were prioritised by the adviser by ranking them from used the most to the least. Additionally, participants reported their ranking of the basketball statistics. This was collected to examine if participants were capable of explicitly learning the adviser's use of the four basketball statistics (H1).

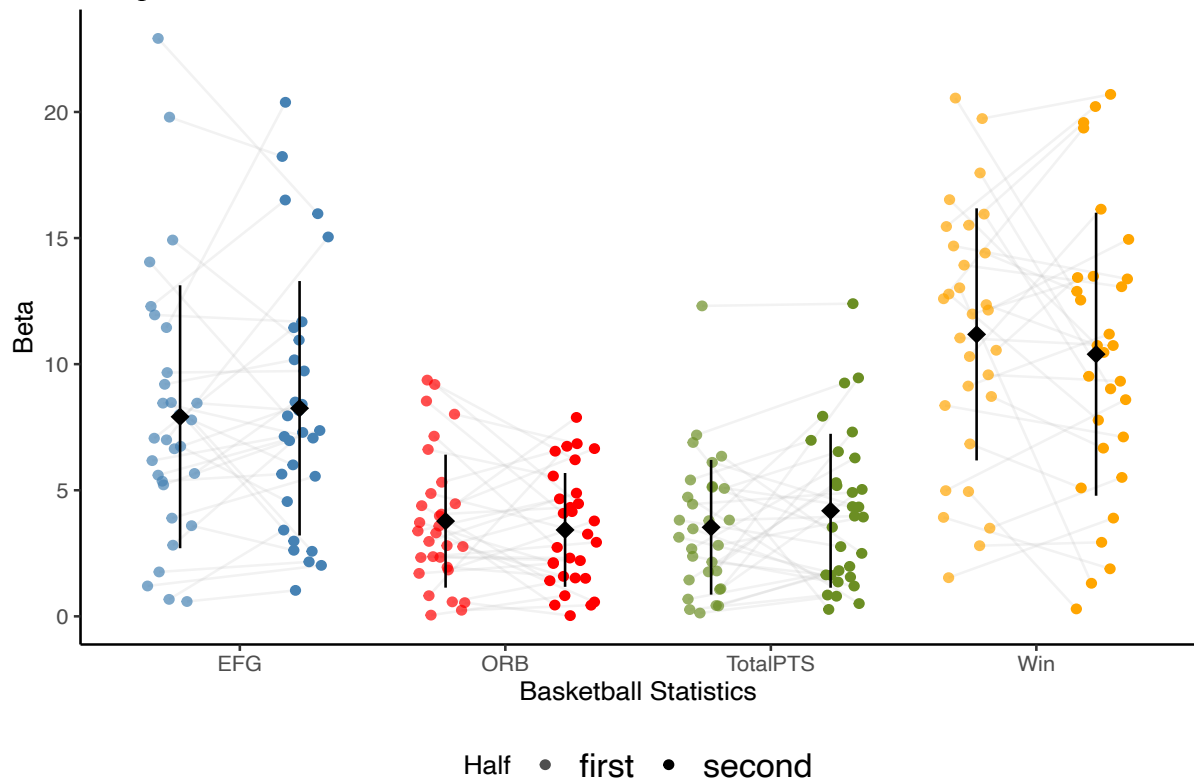
*Estimated Adviser Ranking.* In the experiment, the adviser prioritised the basketball statistics in the following order: Effective Field Goal (eFG%), Offensive Rebound (ORB%), Total Points Scored the Previous Month, Win Percentage This Month. This ranking was compared to the participants' estimation. The average Kendall's rank correlation coefficient across participants was not significantly different from zero ( $M = 0.06$ ,  $SD = 0.52$ ),  $t < 1$ ,  $p = 0.562$ ,  $d = 0.11$ . Therefore, there was no evidence to suggest that participants were able to identify the basketball statistics prioritised by the adviser. Therefore, H1 was not supported. A Bayesian t-test was carried out to investigate evidence for the null hypothesis. There was substantial evidence for the null hypothesis ( $BF_{10} = 0.23$ ,  $\delta = 0.10$ , 95% CrI[-0.24, 0.43]).

*Participants' and Adviser's Ranking.* Participants' ranking of their own and estimated adviser's use of the four basketball statistics was compared. A t-test showed that the Kendall's ranking correlation coefficient was positive on average but not significantly different from zero ( $M = 0.19$ ,  $SD = 0.54$ ),  $t(29) = 1.90$ ,  $p = .067$ ,  $d = 0.35$ . Therefore, there was no overlap between participants' report of their own and the estimated adviser's ranking.

**Use of the Basketball Statistics.** The second approach to testing learning was to investigate how participants' use of the four basketball statistics shifts after repeated interaction with the adviser. In order to test this, the beta coefficients in the first and second half of the trials were compared.

A paired t-test comparing beta coefficient for Effective Field Goal (eFG%) in first half ( $M_{first} = 7.91, SD_{first} = 5.21$ ) and second half ( $M_{second} = 8.25, SD_{second} = 5.04$ ) of the trials found no difference,  $t < 1, p = 0.624, d = 0.09$ . There was also no difference in the use of Offensive Rebound (ORB%) between the first ( $M_{first} = 3.47, SD_{first} = 3.03$ ) and second half ( $M_{second} = 2.81, SD_{second} = 3.01$ ) of the trials,  $t(29) = 1.12, p = .271, d = 0.29$ . This was also the case for the beta values of Points Scored the Previous Month ( $M_{first} = 3.28, SD_{first} = 2.98; M_{second} = 3.53, SD_{second} = 3.81$ ),  $t < 1, p = 0.721, d = -0.07$ . The use of Win Percentage This Month reduced slightly but the difference was not significant ( $M_{first} = 11.18, SD_{first} = 5.00; M_{second} = 10.29, SD_{second} = 5.81$ ),  $t(29) = 1.05, p = .305, d = 0.19$ . The results overall showed no significant difference in the use of the four basketball statistics between the first and second half of the trials and H2 was not confirmed. Four Bayesian t-tests were also conducted to investigate the evidence for no difference in beta values of the four basketball statistics. There was substantial evidence for no difference in EFG from first to second half of the trials ( $BF_{10} = 0.22, \delta = -0.08, 95\% CrI[-0.42, 0.25]$ ). The evidence for no difference in ORB was anecdotal ( $BF_{10} = 0.34, \delta = 0.19, 95\% CrI[-0.15, 0.53]$ ). There was substantial evidence against the difference between Points Scored the Previous Month ( $BF_{10} = 0.21, \delta = -0.06, 95\% CrI[-0.40, 0.27]$ ). This was also the case for Win Percentage This Month ( $BF_{10} = 0.32, \delta = 0.17, 95\% CrI[-0.17, 0.52]$ ).

**Figure 4.3** Participants' use of the four basketball statistics in first and second half of the trials in Experiment 9



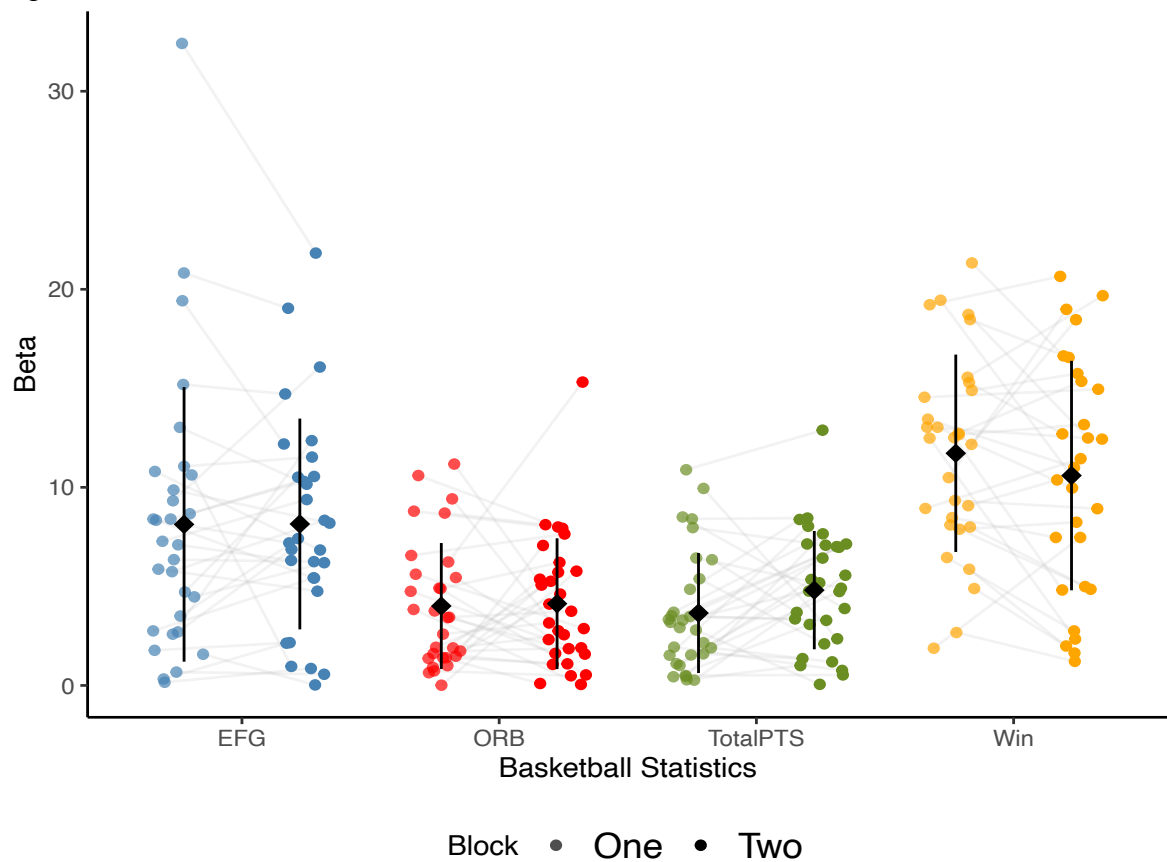
There were numerical differences in the use of the four basketball statistics but none of the differences were significant. One possibility is that participants learned after 24 trials, which was the case in Chapter 3. Therefore, the learning effects were not apparent when comparing first and second half of the trials. Therefore, beta coefficients were calculated for block 1 (first 24 trials) and block 2 (second 24 trials).

Four paired t-test were conducted for each basketball statistics. Comparison of eFG% in block 1 ( $M_{one} = 8.08, SD_{one} = 7.00$ ) and block 2 ( $M_{two} = 8.11, SD_{two} = 5.38$ ) showed no difference,  $t < 1, p = 0.964, d = -0.04$ . No difference between the two blocks were found for ORB% ( $M_{one} = 3.71, SD_{one} = 3.54; M_{two} = 3.27, SD_{two} = 4.18$ ),  $t < 1, p = 0.548, d = 0.20$ , and Points Scored the Previous Month ( $M_{one} = 2.25, SD_{one} = 4.21; M_{two} = 4.00, SD_{two} = 4.04$ ),  $t(29) = -1.73, p = .094, d = -0.07$ . The use of Win Percentage This Month also did not differ by block ( $M_{one} = 11.72, SD_{one} = 4.98; M_{two} = 10.28, SD_{two} = 6.35$ ),  $t(29) =$

1.71,  $p = .098$ ,  $d = 0.19$ . Therefore, there were no significant difference in the use of the four basketball statistics to suggest that participants were learning from adviser confidence. Once again four Bayesian t-tests were also conducted to investigate the evidence for no difference in beta values of the four basketball statistics. There was substantial evidence for no difference in eFG ( $BF_{10} = 0.19$ ,  $\delta = -0.006$ , 95% CrI[-0.34, 0.34]) and ORB ( $BF_{10} = 0.23$ ,  $\delta = 0.10$ , 95% CrI[-0.23, 0.44]) between block 1 and 2. The evidence for no difference in Points Scored the Previous Month ( $BF_{10} = 0.73$ ,  $\delta = -0.28$ , 95% CrI[-0.64, 0.06]) and Win Percentage This Month ( $BF_{10} = 0.32$ ,  $\delta = 0.17$ , 95% CrI[-0.17, 0.52]) was anecdotal.

In addition to comparing at block level, a linear regression was conducted per participant with participants' initial confidence as the dependent variable and the interaction between the four basketball statistics and trial order as the independent variables. The beta values of each interaction term were compared to zero using t-tests, which yielded no significant results (all  $p > 0.05$ ). This demonstrates that participants' use of the four basketball statistics did not change as the trials passed.

**Figure 4.4** Participants' use of the four basketball statistics in block 1 and block 2 in Experiment 9

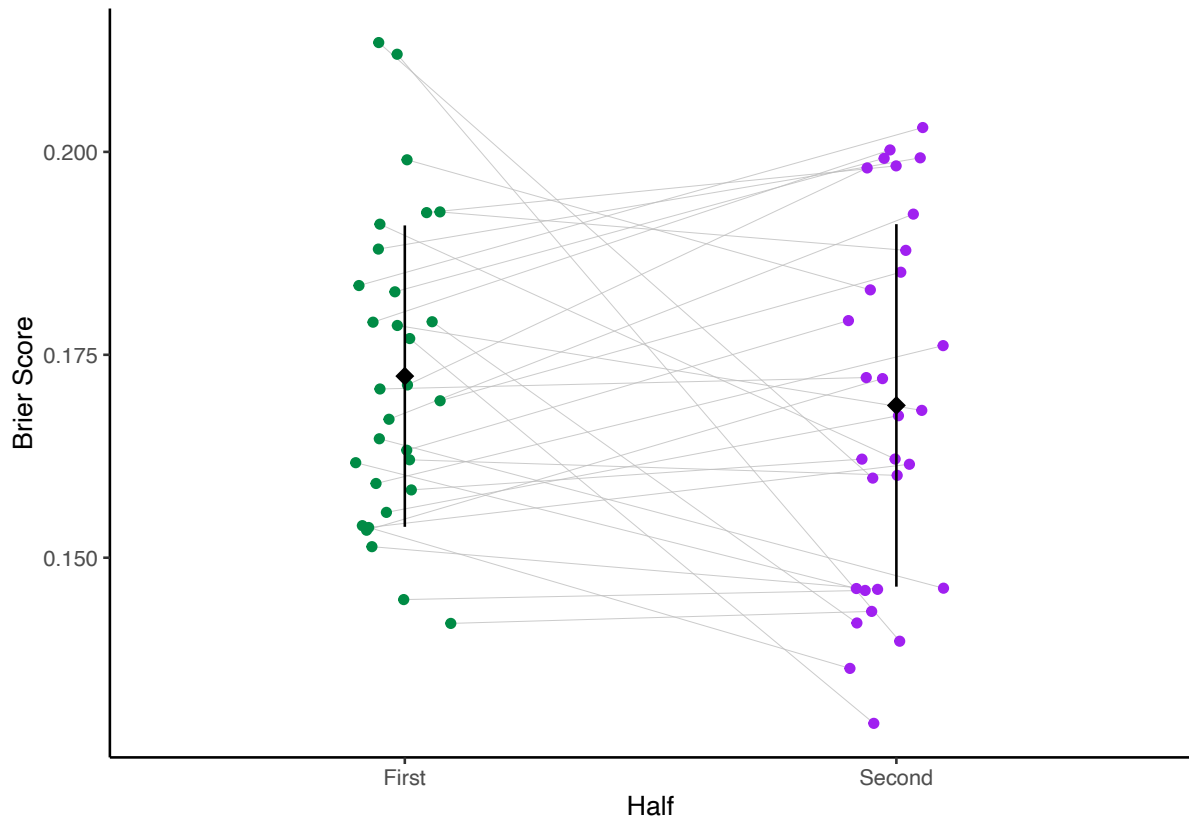


**Accuracy.** Participants' accuracy in the first half (48) and second half (48) of the trials were compared. As visible in Figure 4.5, there was no difference in participants' Brier score between the first half ( $M_{first} = 0.17$ ,  $SD_{first} = 0.02$ ) and the second half ( $M_{second} = 0.17$ ,  $SD_{second} = 0.02$ ) of the trials,  $t < 1$ ,  $p = 0.499$ ,  $d = 0.15$ . In order to better understand the lack of significant results, a Bayesian t-test was conducted to examine the evidence for the null hypothesis. There was substantial evidence for no difference in Brier score from first to second half of the trials ( $BF_{10} = 0.26$ ,  $\delta = 0.13$ , 95% CrI [-0.21, 0.48]).

In addition to the Brier score, participants' binary accuracy in predicting the winning team was also calculated. Participants' initial accuracy from first half ( $M_{first} = 0.77$ ,  $SD_{first} = 0.05$ ) and the second half ( $M_{second} = 0.77$ ,  $SD_{second} = 0.05$ ) of the trials did not differ significantly,  $t < 1$   $p = 0.763$ ,  $d = -0.06$ . Bayesian t-test was once again conducted, which

found substantial evidence for no difference in accuracy from first to second half of the trials ( $BF_{10} = 0.20$ ,  $\delta = -0.06$ , 95% CrI[-0.40, 0.28]).

**Figure 4.5** Participants' Brier score in first and second half of the trials



Participants' accuracy before and after advice was compared to examine if participants became more accurate following advice. Participants' initial ( $M_{initial} = 0.17$ ,  $SD_{initial} = 0.02$ ) and final ( $M_{final} = 0.18$ ,  $SD_{final} = 0.02$ ) Brier score did not significantly differ  $t(29) = -1.72$ ,  $p = .097$ ,  $d = -0.31$ . This was also the case regarding participants' binary accuracy ( $M_{first} = 0.77$ ,  $SD_{first} = 0.04$ ;  $M_{final} = 0.76$ ,  $SD_{final} = 0.03$ ),  $t(29) = 1.28$ ,  $p = .212$ ,  $d = 0.23$ . This limited change in accuracy might be because participants initial accuracy was similar to adviser's accuracy ( $M = 0.75$ ).

**Advice Adoption.** The analyses found no evidence that participants were learning about the validity of the four basketball statistics from adviser confidence. This was puzzling as Chapter 3 found some evidence that participants changed their weighting of the four

basketball statistics after repeated interaction with the adviser. Therefore, further analyses were conducted to ensure that participants were adhering to advice. The adviser influence measure ( $M\Delta_{\text{confidence}}$  for agreement -  $M\Delta_{\text{confidence}}$  for disagreement) from previous chapters could not be calculated since participants' confidence was collected pre-advice and score difference predictions were recorded post-advice. Hence, to measure sensitivity to advice, how frequently participants changed their winning team prediction after agreeing and disagreeing advice was calculated. It was expected that participants would change their predicted winning team more often when an adviser disagrees. Another possible measure of advice adoption was participants' predicted score difference, which was also compared for trials when the adviser agreed versus disagreed with participants' initial answer. If participants were adopting advice, predicted score difference would be smaller after disagreeing advice compared to agreeing advice. Lastly, participants' accuracy pre- and post-advice was examined, as participants' accuracy would increase after adopting advice.

*Prediction Change.* On average, the participants' initial answer and advice agreed about the winning team 82% of the trials. There was a significant difference in the percentage of trials participants changed their answer between agreeing ( $M_{\text{agree}} = 0.00$ ,  $SD_{\text{agree}} = 0.00$ ) and disagreeing ( $M_{\text{disagree}} = 0.42$ ,  $SD_{\text{disagree}} = 0.36$ ) trials,  $t(29) = -6.35$ ,  $p < .001$ ,  $d = -1.16$ . This demonstrates that participants in general rarely changed their answer but were more likely to do so when the adviser disagreed and thus, suggest sensitivity to advice.

*Score Prediction.* Participants' predicted score difference was compared for agreeing and disagreeing trials. There were only a few trials where participants changed their answer (less than 1% of the total trials). Therefore, these trials were excluded from the analysis.

The absolute predicted score difference was used in the analysis, where predicting Home Team would win by 15 points and predicting Away Team would win by 15 points were both recorded as 15. Participants predicted score difference in agreeing ( $M_{agree} = 16.59$ ,  $SD_{agree} = 4.22$ ) and disagreeing ( $M_{disagree} = 9.03$ ,  $SD_{disagree} = 3.55$ ) trials significantly differed,  $t(29) = 10.98$ ,  $p < .001$ ,  $d = 2.00$ . This demonstrates that participants predicted score difference was lower when the adviser disagreed with their initial answer, which in turn indicates that participants were attentive to advice.

### **4.2.3 Discussion**

The aim of the current experiment was to investigate if advisees are able to learn the validity of the external information from adviser confidence. Based on this notion, three hypotheses were proposed. First, it was predicted that advisees would be able to deduce the adviser's weighting of the four basketball statistics (H1). However, the experiment showed no evidence in support of this hypothesis. Participants estimated the adviser's weighting of the four basketball statistics, which was compared to the adviser's actual weighting by calculating a Kendall's rank correlation coefficient. The coefficient was close to zero indicating that participants' estimation did not resemble the adviser's actual ranking. Additionally, participants were asked to state their own ranking of the four basketball statistics. Participants' own weighting and participants' estimated adviser's weighting showed no significant correlation. Second, it was hypothesised that participants would change their own weighting of the basketball statistics after repeated interaction with the adviser (H2). There were some numerical changes in the participants weighting of each basketball statistics, but these differences were not significant and nor were they consistently in the direction that would be expected if they were learning their adviser's strategy. Lastly, participants accuracy was expected to increase following repeated interaction with the adviser (H3). The results showed no support for the hypothesis.

The lack of evidence suggesting that advisees learn from adviser confidence in Experiment 9 was surprising. This deviated from Experiments 6-8 findings, which showed signs of participants inferring the adviser's decision-making strategy and changing their own weighting of the four basketball statistics to resemble the adviser's strategy. For example, participants had difficulties deducing the adviser's ranking of the four basketball statistics (as evident from the low Kendall's rank correlation coefficient). In Chapter 3, the Kendall's rank correlation coefficients were small but positive (e.g., 0.26 in Experiment 6). Additionally, weighting of a couple basketball statistics decreased significantly in some (e.g., Experiment 6, Experiment 8) experiments. These patterns were observed despite the shorter interaction (48 trials) in Experiments 6-8 compared to the current experiment. Furthermore, it was also unanticipated that participants' estimated adviser ranking, and participants' own ranking did not correlate. It was expected that there would be some similarities between participants' own ranking and adviser's ranking. This lack of learning was not because participants were ignoring advice, as participants were changing their predicted winning team more often and anticipated smaller score difference when the adviser disagreed with their answer.

The methodological differences between Experiment 9 and Experiments 6-8 might provide clues for the surprising results in the current experiment. A first possibility is that participants might have felt fatigue due to the longer experimental duration and made less efforts to infer the adviser's decision-making strategy. However, this explanation is unlikely as no difference in the use of the basketball statistics were found when block 1 and 2 were compared, which corresponds to the comparison in Experiments 6-8. Another difference with Experiments 6-8, was that participants were asked to report their confidence during initial decision-making but predicted score difference post-advice.

This change may have prompted participants to focus on the basketball statistics to best predict the score rather than the relationship between adviser confidence and the basketball statistics. Additionally, participants might have found it difficult to process adviser confidence while their main focus was to predict the score difference. Furthermore, the basketball games selected for the current experiment may not have been sufficiently informative. As previously discussed, some discrepancy between the basketball statistics and adviser confidence is essential for participants to learn the validity of the basketball statistics. Basketball games were carefully selected but due to the number of trials in the current experiment, it was difficult to select all games with large discrepancies. For example, in Experiments 6-8, the advisers were highly confident despite ambiguous external evidence ( $M = 81\%$ ). The adviser in the current experiment was not as confident ( $M = 70\%$ ). This may have made it challenging for participants to interpret the adviser's confidence in relation to the basketball statistics.

The lack of findings in Experiment 9 might be due to these methodological variation from Experiments 6-8 but the exact reasons are unclear. No results were found to suggest that participants perceive adviser confidence as a proxy for performance feedback where they could learn about the four basketball statistics. Therefore, Experiment 10 was conducted with methodological changes that better resembled cue learning studies, such as highlighting the benefits of attending to advice. Additionally, a new set of trials were selected that better depicts the relationship between adviser confidence and the basketball statistics.

### 4.3 Experiment 10

In Experiment 9, there were no signs of participants learning from adviser confidence. This may be due to the methodological limitations, but the exact reason is unclear. Therefore, Experiment 10 was conducted with methodological changes that highlights the benefit of learning about the external information from adviser confidence. Additionally, participants did not view advice in all trials. More specifically, participants first predicted the winning basketball team alone with no advice in the first block, which measured participants' baseline performance. In the second and third block, participants interacted with an adviser, giving them a chance to learn the validity of the basketball statistics. In the fourth block, participants once again predicted the game outcomes alone and performance in this block was compared to block 1. Furthermore, participants only interacted with the adviser for 48 trials to enable selection of the most informative trials. Additionally, participants in the top 20% performance were given incentives to motivate participants to pay greater attention to advice. This set-up was expected to provide sufficient time for participants to learn the basketball statistics while minimising fatigue.

#### 4.3.1 Methods

**Participants.** Thirty participants (12 = Female, 18 = Male,  $M_{\text{age}} = 37.6$ ,  $SD_{\text{age}} = 11.21$ ) were recruited through Prolific (<http://prolific.com>) for Experiment 10. In line with the University of Oxford Central University Research Ethics Committee guidelines, informed consent was obtained from all participants before engaging with the experiment. Participants were compensated £9/hour based on the median duration of the experiment.

**Task.** The overall task remained largely similar to Experiment 9, where participants were asked to predict the winning team of a basketball game given four basketball statistics (Effective Field Goal (eFG%), Offensive Rebound (ORB%), Total Points Scored the

Previous Month, Win Percentage This Month). Participants were asked to give their predicted winning team and confidence level on a horizontal scale ranging from 100% Home Team to 100% Away Team, with the mid-point (50%) unselectable. Additionally, there were once again 96 trials split into four blocks. However, some changes were made to the instructions and trial set up to highlight the benefits of learning from adviser confidence. Firstly, there were two types of trials. In the first and last block of the experiment, participants answered the question alone (equivalent to pre-advice answer in Experiments 6-9). On the other hand, blocks two and three followed the same trial set up as Experiments 6-9 (i.e., judge-adviser system). Participants first submitted their initial answer alone and then received advice. While viewing the adviser, participants gave their final predicted winning team and confidence level. This change in experimental set up was to replicate methods used in cue learning studies, where participants performance is measured pre- and post- learning phase. Using two types of trials could accurately measure participants' learning while reducing fatigue. A second modification from Experiment 9 was reporting confidence post-advice (instead of score prediction). Participants in Experiment 9 might have found it difficult to process adviser confidence while their goal was to predict the score difference. Additionally collecting participants' confidence pre- and post-advice could draw more attention to adviser's confidence.

Similar to Experiments 6-9, games where there was a discrepancy between external information ambiguity and adviser confidence were selected, as these trials highlights the relationship between adviser confidence and the four basketball statistics. The trials in Experiment 9 may not have been sufficiently informative. Therefore, new basketball games were selected from the originally collected 168 games, where there was a greater discrepancy between external information ambiguity and adviser confidence than Experiment 9. For example, in the current experiment, when the external information was

ambiguous the adviser was on average ( $M = 79$ ) more confident than in Experiment 9 ( $M = 70$ ). Another key difference was that participants were given accuracy feedback at the end of each block. The accuracy feedback was calculated as the percentage of trials where the winning team prediction was correct (i.e., accuracy in the binary prediction). In all blocks, participants received feedback regarding their own accuracy to help gauge their own performance. In blocks 2 and 3, participants received the adviser's block accuracy as well. This was to emphasise that the adviser was highly accurate which would encourage advice adoption and learning. Indeed, the adviser was more accurate than participants in both block 2 ( $M_{adviser} = 0.91$ ;  $M_{participant} = 0.75$ ) and block 3 ( $M_{adviser} = 0.88$ ;  $M_{participant} = 0.70$ ). In addition, bonus was given to participants with the top 20% accuracy. The bonus was based on participants' Brier score in blocks 1 and 4. More specifically, participants were instructed that their performance would be calculated based on their predicted winning team and its calibration to their confidence reports.

There were additional small methodological changes in Experiment 10. In Experiments 6-9, participants exhibited bias towards Total Points Scored the Previous Month and Win Percentage This Month. This was peculiar as eFG% and ORB% are more informative in predicting the game outcome as they are direct indicators of the game progress. Therefore, additional explanation about the statistics were added to ensure full comprehension of the basketball statistics. Participants were instructed that the eFG% and ORB% are specific to the game whereas Total Points Scored the Previous Month and Win Percentage This Month reflects each team's general performance.

Participants' existing basketball knowledge was also collected because participants with high expertise are less likely to make efforts to deduce the adviser's strategy. Additionally,

people who are basketball fans may have pre-existing conception of the most informative basketball statistics. Therefore, may resist adopting the adviser's use of the four basketball statistics. Participants had varying levels of basketball knowledge (8 = None, 9 = Poor, 9 = Decent, 4 = Good). However, this did not influence participants' advice adoption.

**Adviser.** As in Experiment 9, the adviser was given a name (James) and was described as an expert. However, a different photo from the Chicago Face Database (CFD) (<https://www.chicagofaces.org>) was used to depict the adviser. The adviser was once again represented with a photo of a white male smiling with his mouth closed. A new adviser was selected as he was rated more trustworthy and attractive compared to the previous adviser according to the CFD data. The photo was changed to increase advice adherence as these characteristics have been found to increase advice adoption (Strong, 1968). No other changes were made regarding the adviser. The adviser's answer and confidence were created following the same method as Experiment 6-9, based on the linear regression carried out on the score difference with the four basketball statistics as the independent variables.

**Procedure.** Participants were first instructed that they would predict the winning team of a basketball game and read explanations about the four basketball statistics. Instead of training blocks, participants were given two examples and some guidance on how to predict the winning team. In one example, all four basketball statistics were greater for the Home Team. Therefore, participants were informed that they could be confident that the Home Team would win. In the next example, two basketball statistics were higher for

the Home Team whereas the other two were higher for the Away Team. Participants were told that they should evaluate the information and use their best judgement in such cases.

Before starting the test blocks, participants were instructed that there are four blocks, where in the first and last block, participants would answer the questions alone but receive advice in the second and third block. Additionally, participants were instructed that those with the top 20% accuracy, would receive a bonus, where the precision of their confidence would also be taken into account. Participants' accuracy for the block was provided at the end of each block. After reading all the instruction at their own pace, participants started the first block. Upon completion of the block, participants were asked to rank the four basketball statistics from used the most to the least. On the next page, participants' overall accuracy in the block was given as a percentage. Participants were given a chance to take a break and then proceed.

In the second test block, participants were told that they would complete the next two blocks with the help of an adviser. Participants were then introduced to the adviser and given an example advice. The explanation about the four basketball statistics were once again provided. Upon completion of the second test block, participants were asked to rank the four basketball statistics from used the most to the least once for the adviser and then once again for themselves (i.e., reported two rankings in total). Participants were then provided with their block accuracy and the adviser's block accuracy. Participants were once again given a chance to rest and were provided with the explanation of the four basketball statistics. The third test block was identical to the second test block.

Prior to starting the fourth block, the explanation of the four basketball statistics was once again provided. Participants were also reminded that the precision of their confidence would be reflected in the bonus calculation. Participants then completed the fourth block where the procedure was identical to the first block. Participants predicted the winning team alone and proceeded to the next question after submitting their answer. After completing the block, participants were again asked to rank the four basketball statistics from used the most to the least. Next, participant's block accuracy was provided.

After completing all four blocks, participants answered demographic questions, which were age, gender and basketball knowledge. Basketball knowledge was collected on a scale ranging from 1(= *None*) to 5 (= *Expert*). All participants were debriefed at the end of the experiment. The adviser's competency rating was not collected for this experiment, but it was confirmed in previous experiments that the adviser was perceived to have high expertise.

**Data Analysis.** It was hypothesised that participants would be able to deduce the adviser's weighting of the four basketball statistics from adviser confidence (H1). Therefore, participants were asked to report their estimation of the adviser's use of the four basketball statistics in blocks 2 and 3, for comparison with the adviser's actual use. In order to test H1, Kendall's rank correlation coefficient was calculated for each block. Each rank correlation coefficient was tested against zero to examine if participants could estimate the adviser's ranking of the basketball statistics better than chance level. Participants were also asked to report their use of the four basketball statistics in all blocks. There were two reasons for this. In addition to the hypotheses, I wished to examine if there is an overlap between participants' own ranking and participants' estimated

adviser's ranking of the four basketball statistics in blocks 2 and 3 by calculating a Kendall's rank correlation coefficient. Lastly, participants' own ranking in block 1 and block 4 was compared by analysing each basketball statistics average rank in block 1 and 4. This was to investigate if there was a difference in participants' report of their own weighting of the four basketball statistics between the two blocks.

In Experiment 10, blocks 2 and 3 were designed to help participants learn the validity of the four basketball statistics (i.e., learning phase). Therefore, the main data analyses focused on blocks 1 and 4. To test H2, two linear regressions (one for block 1 and one for block 4) were conducted with participants' confidence (ranging from 0 to 100) as the dependent variable and the difference in the four basketball statistics as the predictors. The linear regressions were conducted separately for each participant. From each linear regression, a set of beta values were derived. Thus, there were two sets of beta values – one from block 1 and another from block 4. The beta values were compared using the same method as Experiment 9, where four paired t-tests were conducted (one for each basketball statistics). Lastly, participants' Brier score and accuracy in predicting the winning team were compared between blocks 1 and 4, using two paired t-test (one for each measure). This was to investigate if participants became more accurate after interacting with the adviser (H3).

#### **4.3.2 Results**

**Ranking Correlation.** Experiment 10 was conducted to examine if advisees can learn the predictiveness of external information from adviser confidence. It was hypothesised that participants would be able to deduce the adviser's weighting of the four basketball statistics (H1). This hypothesis was to test participants' explicit learning. In addition to

the hypothesis, participants also reported their ranking of the four basketball statistics after each block. Participants own ranking was compared to participants' estimated adviser's ranking in block 2 and 3. Participants' own ranking in blocks 1 and 4 were also analysed.

*Estimated Adviser's Ranking.* Hypothesis 1 was tested by inspecting participants' accuracy in estimating the adviser's ranking of the four basketball statistics. A t-test comparing the Kendall's rank correlation coefficient to zero found significant differences in both block 2 ( $M_{two} = 0.26$ ,  $SD_{two} = 0.53$ ),  $t(29) = 2.60$ ,  $p = .014$ ,  $d = 0.48$ , and block 3 ( $M_{three} = 0.43$ ,  $SD_{three} = 0.52$ ),  $t(29) = 4.57$ ,  $p < .001$ ,  $d = 0.83$ . This suggests that participants were able to partially deduce the basketball statistics prioritised by the adviser. The difference between the two blocks were not significant,  $t(29) = -1.89$ ,  $p = .070$ ,  $d = -0.34$ . However, there was a still a numerical increase in the Kendall's rank correlation coefficient which hints that participants became slightly more accurate in estimating the adviser's weighting of the four basketball statistics with greater interaction. Bayesian t-tests also showed similar patterns. There was substantial evidence suggesting that the Kendall's rank correlation coefficient in block 2 was different from zero ( $BF_{10} = 3.32$ ,  $\delta = 0.44$ , 95% CrI[0.08, 0.81]). There was also decisive evidence suggesting that Kendall's rank correlation coefficient in block 3 was significantly different from zero ( $BF_{10} = 307.43$ ,  $\delta = 0.78$ , 95% CrI[0.37, 1.19]). There was anecdotal evidence against the difference in rank correlation coefficient between the two blocks ( $BF_{10} = 0.93$ ,  $\delta = -0.31$ , 95% CrI[-0.68, 0.04]).

*Participants' and Estimated Adviser's Ranking.* The Kendall's rank correlation was calculated between participants' own and participants' estimated adviser's use of the four

basketball statistics. This was to test if there are any overlaps between the two rankings. The Kendall's rank correlation coefficient in block 2 was significantly different from zero ( $M_{two} = 0.58$ ,  $SD_{two} = 0.44$ ),  $t(29) = 7.24$ ,  $p < .001$ ,  $d = 1.32$ . This was also the case in block 3 ( $M_{three} = 0.58$ ,  $SD_{three} = 0.53$ ),  $t(29) = 5.95$ ,  $p < .001$ ,  $d = 1.09$ . These results demonstrate that there was some overlap between participants' reported own ranking and participants' estimation of the adviser's ranking. There was no significant difference between the two blocks,  $t < 1$ ,  $p = 1$ ,  $d = 0$ .

*Participants' Stated Ranking.* Participants' own ranking of the four basketball statistics in block 1 and block 4 was compared to examine if participants changed their decision-making strategy. Participants' ranking for each basketball statistic was compared between block 1 and 4 through four paired t-tests (one for each statistic). Participants' reported ranking of Effective Field Goal (eFG%) showed no significant difference between block 1 ( $M_{one} = 2.33$ ,  $SD_{one} = 1.06$ ) and block 4 ( $M_{four} = 2.00$ ,  $SD_{four} = 1.11$ ),  $t(29) = 1.67$ ,  $p = .106$ ,  $d = 0.30$ . However, there was a numerical reduction, indicating that participants ranked eFG% higher (i.e., assigned greater weight) in block 4. Participants ranked Offensive Rebound (ORB%) significantly higher in block 4 ( $M_{four} = 2.43$ ,  $SD_{four} = 1.07$ ) compared to block 1 ( $M_{one} = 2.93$ ,  $SD_{one} = 1.17$ ),  $t(29) = 2.06$ ,  $p = .049$ ,  $d = 0.38$ . The ranking of Total Points Scored the Previous Month increased, indicating that participants reported using the statistic less but the difference was not significant ( $M_{one} = 2.40$ ,  $SD_{one} = 1.35$ ;  $M_{four} = 2.77$ ,  $SD_{four} = 1.10$ ),  $t(29) = -1.78$ ,  $p = .086$ ,  $d = -0.32$ . On the other hand, the ranking of Win Percentage This Month significantly increased from block 1 ( $M_{one} = 2.33$ ,  $SD_{one} = 0.76$ ) to block 4 ( $M_{four} = 2.80$ ,  $SD_{four} = 1.06$ ),  $t(29) = -2.45$ ,  $p = .020$ ,  $d = -0.45$ . These findings demonstrate that there were some changes in participants' reported use of the four basketball statistics after interacting with the adviser. Additionally,

participants ranked the statistics prioritised by the adviser (eFG%, ORB%) higher, while lowering the ranks of the statistics used less (Total Points, Win Percentage) by the adviser. This suggests some explicit learning from adviser confidence.

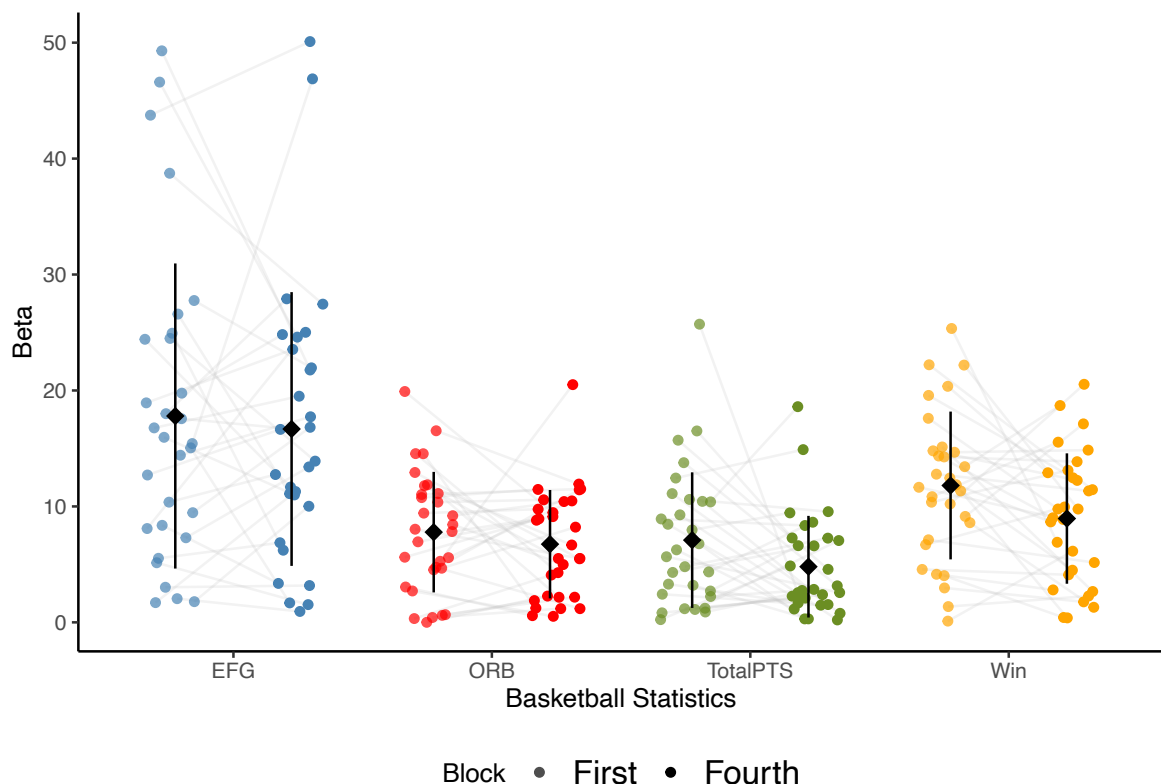
***Use of Basketball Statistics.*** A second test of learning was to examine if participants' weighting of the four basketball statistics changed from block 1 to block 4 (H2). In contrast to H1 which tested explicit learning, H2 tested if participants implicitly learned from adviser confidence and changed their behaviour accordingly. In order to test H2, a set of beta values were derived in block 1 and 4 from linear regressions assessing the relationship between participants confidence and the four basketball statistics. Participants' use of the four basketball statistics was compared between block 1 and 4. Four separate paired t-tests were conducted to compare the beta values - once for each basketball statistics. The analysis of Effective Field Goal (eFG%) showed no significant difference between block 1 ( $M_{one} = 17.25, SD_{one} = 13.89$ ) and block 4 ( $M_{four} = 16.51, SD_{four} = 12.04$ ),  $t < 1, p = .741, d = 0.06$ . The analysis of Offensive Rebound (ORB%) also showed no significant difference between the two blocks ( $M_{one} = 7.56, SD_{one} = 5.53; M_{four} = 6.39, SD_{four} = 5.17$ ),  $t(29) = -1.08, p = .288, d = 0.20$ . There was also no significant difference in the use of Total Points Scored the Previous Month ( $M_{one} = 6.77, SD_{one} = 6.23; M_{four} = 4.36, SD_{four} = 4.83$ ),  $t(29) = -1.82, p = .079, d = 0.33$ , and Win Percentage This Month ( $M_{one} = 10.88, SD_{one} = 7.89; M_{four} = 8.63, SD_{four} = 6.12$ ),  $t(29) = -1.68, p = .105, d = 0.31$ . Although the differences were not significant, participants' use of all four basketball statistics numerically decreased (Figure 4.6).

The minimal change in participants' use of the four basketball statistics was puzzling. Therefore, Bayesian t-tests were conducted to better understand the results. There was

substantial evidence for no difference in EFG from first to second half of the trials ( $BF_{10} = 0.20$ ,  $\delta = 0.05$ , 95% CrI[-0.29, 0.39]). The evidence for no difference in ORB was substantial ( $BF_{10} = 0.33$ ,  $\delta = 0.18$ , 95% CrI[-0.16, 0.52]). There was anecdotal evidence against the difference between Points Scored the Previous Month ( $BF_{10} = 0.83$ ,  $\delta = 0.30$ , 95% CrI[-0.05, 0.66]). This was also the case for Win Percentage This Month ( $BF_{10} = 0.68$ ,  $\delta = 0.28$ , 95% CrI[-0.07, 0.63]).

Trial effect was investigated by running a linear regression per participant with participants' initial confidence as the dependent variable and the interaction between the four basketball statistics and trial order as the independent variable. The beta values of each interaction term were compared against zero using t-tests (conducted four t-tests in total). The t-tests showed no difference from zero except for Win Percentage This Month ( $M_{Win} = 4.36$ ,  $SD_{Win} = 4.83$ ),  $t(29) = 2.20$ ,  $p = .036$ ,  $d = 0.40$ .

**Figure 4.6** Participants' use of the four basketball statistics in block 1 and 4 in Experiment 10



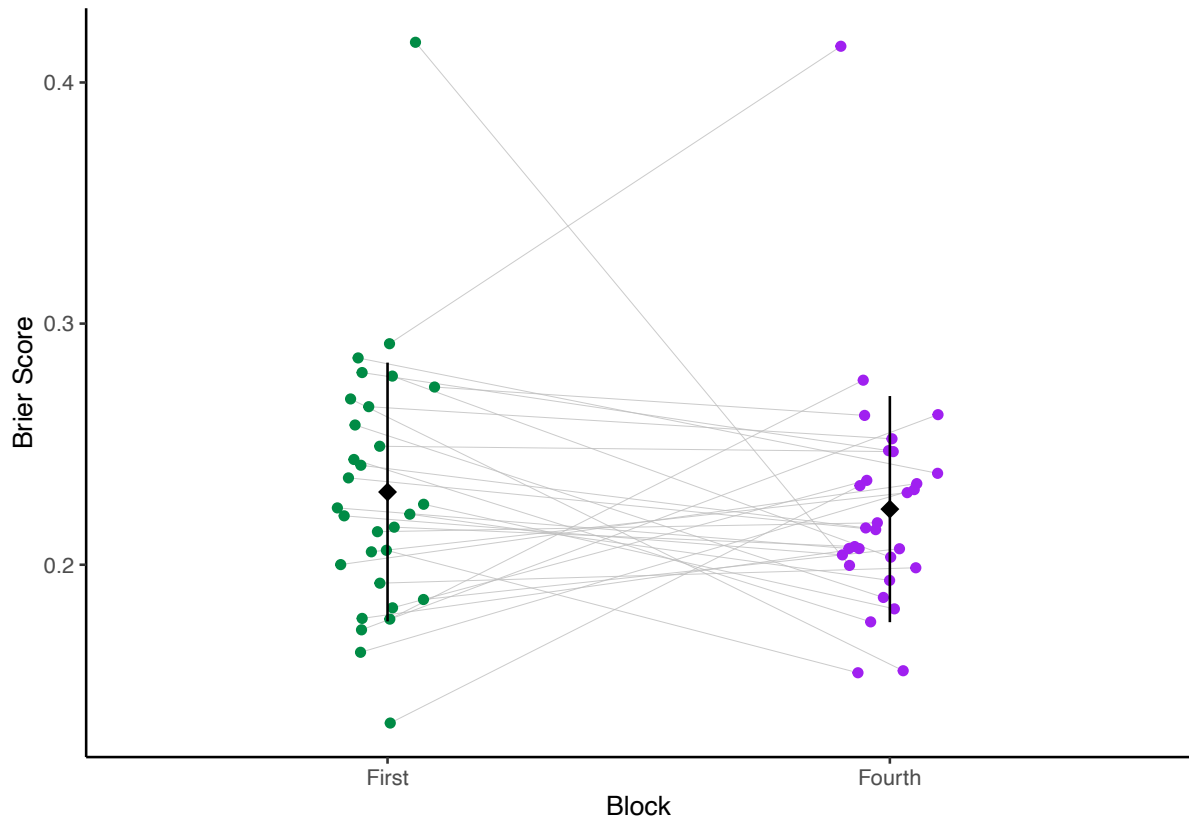
**Accuracy.** It was also predicted that participants' accuracy would increase as they learned about the four basketball statistics from adviser confidence (H3). Participants' Brier score in block 1 and 4 was compared to assess if there was an improvement in participants' accuracy-confidence calibration. There was no significant difference in participants' Brier score between block 1 ( $M_{one} = 0.23$ ,  $SD_{one} = 0.05$ ) and block 4 ( $M_{four} = 0.22$ ,  $SD_{four} = 0.05$ ),  $t < 1$ ,  $p = .577$ ,  $d = 0.10$ . Therefore, no evidence was found to suggest that participants' accuracy increased after repeated interaction with the adviser and there were individual variations in the Brier score trend (Figure 4.7). Therefore, H3 was not supported. A Bayesian t-test also found substantial evidence for the absence of difference in Brier score between the two blocks ( $BF_{10} = 0.23$ ,  $\delta = 0.09$ , 95% CrI[-0.24, 0.43]).

Although the main measure of accuracy was the Brier score, participants' binary (i.e., predicted winning team) accuracy was also examined. Participants' accuracy in block 1 ( $M_{one} = 0.68$ ,  $SD_{one} = 0.07$ ) and block 4 ( $M_{four} = 0.67$ ,  $SD_{four} = 0.07$ ) did not differ significantly,  $t < 1$ ,  $p = .634$ ,  $d = 0.09$ . Bayesian t-test also found substantial evidence to suggest no difference in accuracy between the two blocks ( $BF_{10} = 0.22$ ,  $\delta = 0.08$ , 95% CrI[-0.26, 0.42]).

Additionally, participants' accuracy pre- and post-advice in Blocks 2 and 3 were examined to investigate whether advice-taking increased accuracy. Participants' initial ( $M_{initial} = 0.19$ ,  $SD_{initial} = 0.05$ ) and final ( $M_{final} = 0.15$ ,  $SD_{final} = 0.04$ ) Brier score significantly differed  $t(29) = 5.77$ ,  $p < .001$ ,  $d = 1.05$ . Similar patterns were observed regarding participants' binary accuracy score, where participants' final accuracy score

( $M_{final} = 0.81$ ,  $SD_{final} = 0.09$ ) was significantly higher than participants' initial ( $M_{initial} = 0.73$ ,  $SD_{initial} = 0.11$ ) accuracy score,  $t(29) = -4.59$ ,  $p < .001$ ,  $d = -0.84$ .

**Figure 4.7** Participants' Brier score in block 1 and 4 in Experiment 10



**Adviser Influence.** Experiment 10 results showed limited evidence of learning from adviser confidence. In order to ensure that participants were adhering to advice, adviser influence was calculated. Adviser influence was calculated following the same method as Experiments 1-5, which was the difference in mean change of confidence in agreeing and disagreeing trials ( $M\Delta_{\text{confidence}}$  for agreement -  $M\Delta_{\text{confidence}}$  for disagreement). There was a significant difference in participants' confidence change in agreeing ( $M_{\text{agree}} = 2.22$ ,  $SD_{\text{agree}} = 3.40$ ) compared to disagreeing ( $M_{\text{disagree}} = -15.12$ ,  $SD_{\text{disagree}} = 13.51$ ) trials,  $t(29) = 7.16$ ,  $p < .001$ ,  $d = 1.31$ . This demonstrates that participants were adopting advice and were sensitive to whether advice agreed or disagreed.

### 4.3.3 Discussion

In Experiment 9, participants showed no signs of learning from adviser confidence. There were possible reasons for the lack of findings, such as the trial set up (e.g., score prediction post-advice) and the informativeness of the basketball games selected as stimuli. Therefore, Experiment 10 was conducted to address these methodological limitations while emphasising the benefits of adhering to advice (e.g., providing accuracy feedback and incentives for prediction accuracy).

Participants explicit and implicit learning of the external information validity from adviser confidence was tested. In terms of explicit learning, it was predicted that participants would be able to deduce the adviser's ranking of the four basketball statistics (H1). In Experiment 10, participants were better able to infer the adviser's use of the four basketball statistics compared to Experiment 9. The Kendall's ranking correlation coefficient was significantly greater than zero in both block 2 and 3. Additionally, although the difference was not significant, participants were numerically better able to estimate the adviser's ranking in block 3. This suggests that although not perfectly, advisees can construe the adviser's decision-making strategy from adviser's decisions and associated confidence.

Furthermore, there were some correlations between participants' estimated adviser's ranking and participants' own ranking of the four basketball statistics. In addition, there was some change in participants' reported use of the four basketball statistics from block 1 to block 4. In particular, ranking of basketball statistics used most by the adviser either improved numerically (eFG%) or significantly (ORB%). On the other hand, participants reported using Total Points Scored the Previous Month (numerically) and Win Percentage

This Month (significantly) less in block 4 than block 1, which were prioritised less by the adviser.

Although participants stated ranking of the four basketball statistics changed from block 1 to block 4, this was not reflected in participants' decision-making. Hypothesis 2 tested participants' implicit learning of the external information by examining their use of the four basketball statistics from block 1 to 4. The analysis found no significant difference in the weighting of the statistics. The beta coefficient numerically decreased the most for Total Points Scored the Previous Month and Win Percentage This Month, which were used less by the adviser. It is possible that participants learned that these statistics were used less by the adviser and thus, placed less weight on these statistics in block 4. However, all beta coefficients numerically decreased, and advice is unlikely to have played a role.

One difference from Experiments 6-9 was that participants prioritised Effective Field Goal (eFG%) prior to advice-taking. This might be because in Experiment 10, it was specified that eFG% relates to the current game status, which led people to over rely on eFG% even before interacting with the adviser. Therefore, it was improbable that the use of eFG% would increase further despite interaction with the adviser. However, surprisingly, participants' use of Offensive Rebound (ORB%) remained low despite being guided that it is also about the current game. This also deviates from participants' own reported ranking results as participants ranked ORB% higher in block 4 than in block 1. Additionally, it was hypothesised that participants would become more accurate as they learn the validity of the four basketball statistics (H3). Participants accuracy did not

improve in block 4 compared to block 1 (i.e., no change in Brier score nor binary accuracy).

In sum, the experiment found weak signs of explicit learning. The ranking results together suggest that participants were able to somewhat learn the adviser's ranking of the four basketball statistics, which slightly improved with repeated interaction. Furthermore, participants reported shifting their use of the four basketball statistics, which somewhat resembled the adviser's true ranking. On the other hand, participants were only partially able to estimate the adviser's ranking. Additionally, there were no signs of implicit learning from block 1 to 4, which was reflected in participants actual use of the four basketball statistics and their ability to predict the winning team. The lack of change in the use of the four basketball statistics is puzzling. Although, the effect was small in Experiment 6 and 8, there was still a shift. Furthermore, the effects were present even though the benefits of learning from adviser confidence were not emphasised. In the current experiment, the benefits of copying an adviser were clear (adviser was more accurate than participants). Additionally, the experimental set-up resembled cue learning studies where participants exhibited learning despite probabilistic feedback.

The lack of significant findings compared to Experiments 6-8 might be due to some methodological differences. First, although the games in Experiment 10 were more informative than Experiment 9, the games selected for Experiment 10 might still not have been sufficiently informative compared to Experiments 6-8. For example, the adviser was on average more confident despite ambiguous external information in Experiment 10 ( $M = 79\%$ ) compared to Experiment 9 ( $M = 70\%$ ) but still less confident than Experiments 6-8 ( $M = 81\%$ ). It is possible that learning from adviser confidence only occurs when

there is a large discrepancy between adviser confidence and external information ambiguity. Second possible explanation is that participants might have focused on finding the correct winning team and confidence rather than the relationship between the external information and adviser confidence. Participants were instructed that there is a bonus for accuracy-confidence calibration. Therefore, participants might have simply focused on how to maximise their trial decision accuracy-confidence calibration rather than improving in their general decision-making strategy.

A more general possible explanation for the lack of findings is that advisees did not perceive advice as a proxy as accuracy feedback. In cue learning research, participants receive partial feedback but are aware that they will receive objective feedback. In this experiment, participants may have viewed the adviser to have high levels of expertise but still inaccurate. Also, participants had difficulties deducing the adviser's use of the four basketball statistics. Although there were improvements compared to Experiment 9, participants were still far from accurately guessing the adviser's ranking (i.e., less than one). This difficulty in inferring the adviser's decision-making strategy coupled with lack of trust in advice might have led participants to disregard adviser's use and maintain their own use of the basketball statistics.

#### **4.4 General Discussion**

In this chapter, two experiments were conducted to investigate if advisees use adviser confidence to learn about the external information (e.g., validity of basketball statistics). In Chapter 3, there were signs of participants adjusting their weighting of the four basketball statistics to better resemble the advisers' weighting after repeated interaction with the adviser. Therefore, Experiments 9-10, were designed to further investigate this

learning effect of adviser confidence with methodological changes to maximise learning (e.g., via longer interaction). Despite these methodological modifications, the experimental results found limited evidence that participants could learn the adviser's use of the external information. In Experiment 9, there were no signs of explicit learning (estimating the adviser's ranking of the four basketball statistics) nor implicit learning (participants change their use of the basketball statistics). One possible reason was that the basketball games selected for Experiment 9 were not sufficiently informative. Therefore, a new set of games were used in Experiment 10. This led to some improvements, where participants were better able to identify the adviser's ranking of the basketball statistics which also ameliorated after repeated interaction with the adviser. In addition, participants reported changing their ranking of the basketball statistics from block 1 to 4, which partially resembled the adviser's ranking. However, there were still no signs of implicit learning as participants actual weighting of the four basketball statistics from showed no significant difference pre- and post- advice-taking.

The findings were unexpected considering previous research on cue learning. Previous literature has shown that people are capable of learning under various types of feedback such as partial or conditional feedback (e.g., feedback only when correct) (De Klerk, 1972; Hendrix & Dudycha, 1981; Stewart et al., 2012), and probabilistic feedback (Little & Lewandowsky, 2009). Moreover, providing probabilistic feedback helped people better learn the relationship amongst cues and even detect cue relationships that are irrelevant to the decision-making process (Little & Lewandowsky, 2009). The adviser's answer and confidence are not a definitive form of feedback, but the adviser was mostly correct and thus, mimicking probabilistic feedback. Moreover, Chapter 3 results suggested that advisees infer the advisers' decision-making strategy from adviser confidence and adjust

their own use of the external information. For example, in Experiment 6, there was a significant decrease in the use of the basketball statistics used less by the advisers. Therefore, it was expected that participants would be able to learn the validity of the basketball statistics in the current experiments.

The lack of learning in this chapter could be explained by how advice was construed. For instance, participants may not have perceived advice as a proxy for feedback. In previous cue learning studies, participants were aware that the provided feedback was objectively accurate and did not have to question the authenticity of the feedback. Even under probabilistic feedback, participants were told in advance that the feedback is not always accurate and were given a range of accuracy participants should aim for. Thus, although the feedback was random, participants were aware of the degree of randomness which partially eliminates uncertainty. Researchers have found that not providing such information could make participants frustrated and sometimes abandon efforts to evaluate cue validity (Kruschke & Johansen, 1999). However, in the current experiments, participants were not provided with the adviser's exact accuracy. The adviser's general expertise was provided, which described him as mostly accurate, but this may not have been sufficient to eradicate all mistrust in the adviser. Therefore, participants might have perceived the adviser's confidence as another source of uncertainty where they need to identify its connection to the basketball statistics (e.g., strength of correlation). The adviser's accuracy information was not provided prior to the experiment to mimic real-world advice-taking contexts. Advisees often do not have access to the precise accuracy of an adviser but simply their general level of expertise. Therefore, only the description of the adviser was provided in the experiments.

Additionally, previous research has found that people's own confidence guides cue learning (Guggenmos et al., 2016). This suggests that how people interpret external information is also contingent on one's own confidence, which also influences the processing of adviser confidence. Indeed, people generally prefer to access information that confirms their stance and underweight disproving evidence (Kaanders et al., 2022). Numerous studies have also found that advisees process advice in relation to their own answer and confidence. For example, people tend to ignore advice that is distant from their own answer (Newell et al., 2009) and perceive advisers who disagree with them as less competent (Pescetelli & Yeung, 2021). On the other hand, advisees become overconfident regardless of the objective accuracy of their answer when an adviser agrees with them (Soll et al., 2021). This pattern was observed even with multiple advisers. Advisees become more confident as long as one adviser agrees with them and ignore the second adviser who disagrees with them (Molleman et al., 2020). Therefore, participants may have processed adviser confidence in a similar manner in the current experiments. For example, in Experiment 9, participants rarely changed their initially predicted winning team (less than 1% of the trials) despite disagreement from the adviser. This underweighting of advice (i.e., egocentric discounting) may be elicited by different factors. One possible reason is that advisees have access to their own reasoning but are deprived of the adviser's thought process (Yaniv & Kleinberger, 2000). In support of this, research found that increased access to the adviser's reasoning reduced egocentric discounting (Van Swol, 2011). In the current experiments, participants might have had difficulties understanding why the adviser favoured one basketball statistic over another and decided to not adopt the adviser's use of the four basketball statistics.

The above-mentioned studies provide hints as to why there was a lack of learning in Experiments 9-10 compared to previous literature. However, it is still unclear why some learning was observed in Experiments 6-8 but not in the current chapter. Methodological differences might partially explain the lack of findings. Firstly, there was only one adviser in Experiments 9-10 whereas there were two advisers in Experiments 6-8. The use of a novice and an expert adviser in Experiments 6-8 might have led participants to more carefully evaluate the adviser's ability compared to Experiments 9-10. For instance, participants might have perceived the novice as a benchmark to compare the expert's performance. Additionally, participants might have heeded to external information more in order to determine the advisers' ability, which might have led participants to better detect the relationship between adviser confidence and external information. In contrast, as there was only one adviser in Experiments 9-10, participants might have not felt the need to evaluate the adviser and simply focused on the degree of advice adoption. A second difference might have been that the basketball games selected for Experiments 9-10 were not as informative as games in Experiments 6-8. In Experiments 6-8, the discrepancy between external information ambiguity and adviser confidence was greater. For example, when the external information was ambiguous (e.g., only two statistics high for Home Team) the advisers were at least 77% confident with a mean confidence of 81%. In contrast, in Experiments 9-10, the adviser was on average 79% confident and their minimum confidence was 69%. Participants might have perceived adviser confidence as less surprising and spent less time and effort to understand unexpected adviser's confidence. On a similar note, the lack of trials where there were no discrepancies between adviser confidence and external information ambiguity (i.e., filler trials) might also explain the difference in findings between Chapter 3 and 4. The inclusion of filler trials might have made trials where there were discrepancies (i.e., test trials) more

unexpected. This surprise may have encouraged participants to carefully examine these trials. The absence of filler trials in Experiments 9-10, might have led participants to believe that the adviser was inherently highly confident/uncertain despite ambiguous/unambiguous external information, and thus paid limited attention.

The results from Chapter 3 and 4 taken together suggest that learning from adviser confidence only occurs under specific conditions and there is only a weak effect. One possible explanation of the results might be due to the individual difference in learning under limited feedback. For example, Herzog and Fahle (1997) found large individual variations in reinforcement learning with no feedback. Research in cue learning under limited feedback has also found individual variations (Guggenmos et al., 2016; Jozefowicz, 2021). In addition, people show wide disparity in susceptibility to social information (Toelch et al., 2014). For instance, some participants might have been unwilling to adopt the adviser's decision-making strategy even before interacting with the adviser. Second possible reason is that people generally struggle to learn complex relationships. More specifically, studies have found that people have difficulties learning non-linear cue relationships (Klayman, 1984). For example, Newell and colleagues (2009) suggest that people often perceive cues as additive (e.g., tally number of cues suggesting outcome A) and assume an independent linear relationship between each cue and criterion. However, cues may have multiplicative rather than simply additive relationships. This suggests that people generally prefer to use simple strategies in decision-making. Therefore, in the current experiments, participants might have tried to find the most informative basketball statistics rather than learning the relationship among the four basketball statistics. This may also explain why only the use of a couple of basketball statistics changed in Experiment 6 and 8. Participants might not have learned the adviser's

full decision-making strategy but simply pursued to identify the most diagnostic basketball statistics and changed their weighting accordingly. This suggests that it may be possible for advisees to learn simple relationships between advice and a couple external information. This was not investigated in the current experiments as the purpose was to examine if learning from adviser confidence can occur in complex environments that closely resembles the real-world. Yet, understanding under which circumstances learning occurs from advice would be an interesting avenue for future research.

Another possible reason for the lack of findings may be due to the individual's pre-conception of the task. In Experiment 10, some participants reported having some knowledge in basketball. This did not influence participants advice adoption but participants with greater knowledge may still be adamant in changing their decision-making strategy. Even participants with no knowledge in basketball might have developed preferences for a certain basketball statistic during the experiment. Therefore, participants may have been biased about which basketball statistics are relevant for decision-making. This was evident from the overweighting of Win Percentage This Month in previous experiments. Thus, advice might not have been sufficient to change advisees' existing perception regarding the validity of the basketball statistics. Indeed, research in cue learning also found that established cue validity is difficult to change through feedback (Murphy & Wisniewski, 1989). Another possibility is that advisees might not wish to adopt adviser's decision-making strategy regardless of whether they could deduce the adviser's strategy. Research in social information suggest that people do not simply copy others' behaviour but modify and improve them (Derex et al., 2015). This tendency is unlikely to explain the results in Experiments 6-10, as there were overall

only small changes in the weighting of the basketball statistic. However, future research could explore if advisees simply adopt or modify adviser's decision-making strategy.

#### **4.4.1 Conclusion**

Whether advisees could learn the validity of the external information from adviser confidence was tested in two experiments. It was predicted that adviser confidence would play a similar role as probabilistic feedback in cue learning and result in learning. However, no consistent signs of learning of external information were observed in the current experiments. There are few possible explanations for the lack of findings. In light of the methodological differences with Experiments 6-8, it is possible that learning only occurs under specific conditions. For example, when the relationship between external information and adviser confidence is distinct. Additionally, previous literature suggests that people have difficulties learning complicated relationships (e.g., non-linear relationships between cues and outcomes). Learning about the external information from adviser confidence may only occur if the relationship between adviser confidence and external information is simple, which could be explored in future research.

# Chapter 5: General Discussion

---

## 5.1 Summary of Experiments

The aim of this thesis was to examine if advisees flexibly interpret adviser confidence. Previous research has focused on the idea that adviser confidence is used as cue to their competence, as exemplified in the confidence heuristic (Price & Stone, 2004). The research in this thesis extended this idea to explore the inferences advisees make about the adviser and how advice is processed in relation to the advice-taking context. Prior to delving into the experiments, Chapter 1 first reviewed research in the advice-taking literature. Past research suggest that people generally seek advice to gain further information and improve their decision-accuracy. Despite this motive, people often overweight their own opinion and do not use advice to its full potential (i.e., egocentric discounting). The degree of egocentric discounting is influenced by various factors, but in the experimental chapters, advisers' expertise and confidence were further explored. In the case of adviser's expertise, people naturally perceive greater expertise as a sign of higher competence and thus, increase advice adoption (Bonnacio & Dalal, 2010; Sniezek et al., 2004; Jungermann & Fischer, 2014). Similarly, advisers who express high levels of confidence are more influential (Sniezek & Buckley, 1995; Sniezek & Van Swol 2001; Van Swol, 2009; Van Swol & Sniezek, 2001; Wang et al., 2025; Yaniv 1997). As previously mentioned, this is because people assume that confident advisers have greater information and are more competent (i.e., confidence heuristic) (Price & Stone 2004; Thomas & McFadyen, 1995). However, adviser confidence and accuracy are not always well-calibrated, as people often overestimate their accuracy (Van Swol & Sniezek 2005; Sah et al., 2013, Klayman et al., 1999; Lichtenstein et al., 1982). In addition to assumptions about accuracy, people make other inferences about an adviser's confidence,

such as their trustworthiness (Sniezek & Van Swol, 2001) and the level of commitment to their advice (Vullioud et al., 2017). These findings demonstrate that advisees process adviser confidence in a complex manner.

Therefore, the experimental chapters explored how adviser confidence is processed in different contexts. Chapter 2 investigated how advisees construe the relationship between adviser's linguistic expressions of confidence and different types of contextual uncertainty. In particular, two types of uncertainties (epistemic vs. aleatory) were investigated (Ülkümen et al., 2016). Epistemic uncertainty refers to uncertainties that could be resolved through acquisition of further information or knowledge. On the other hand, aleatory uncertainty arises due to the inherent randomness in the environment (e.g., dice roll). Ülkümen and colleagues (2016) found that people tend use confidence terms (e.g., certain, sure, confident) to describe epistemic uncertainty whereas likelihood terms (e.g., probability, chance, likelihood) are preferred to illustrate aleatory uncertainty. Whether people prefer advisers to use context appropriate linguistic expressions was examined in five experiments. The experiments adopted a mixed-design, where all participants received advice from two advisers—one of whom used confidence terms and another who used likelihood terms. One group of participants performed tasks that were designed to have high levels of epistemic uncertainty (e.g., historic dates, total area) while the other group engaged in a task with high levels of aleatory uncertainty (e.g., temperature prediction, coin-toss). Only one experiment found a robust difference in perceived adviser competence and adviser influence across and within the two of advice-taking contexts (Experiment 2). There were numerical differences in the degree of advice adoption (Experiment 3) and perceived adviser competence (Experiment 4) within one task in the expected direction. However, Experiment 5, which used identical methods as

Experiment 2, failed to replicate the findings. Furthermore, in Experiment 2 the effect sizes were small to medium, and the effect sizes in Experiments 1 and 3-5 were close to zero ( $< 0.01$ ). This suggest that increasing the number of participants is unlikely to find different results from the current findings. Therefore, the experimental findings taken all together suggest that people pay limited attention to which linguistic expressions are used in advice and are likely to only exhibit a small preference for advisers who use context appropriate linguistic expressions. Nonetheless, the experiments found that people consistently distinguish between epistemic and aleatory uncertainty. For example, the tasks with knowable answers (e.g., estimating historic dates, comparing total area) were perceived as having high levels of epistemic uncertainty but low levels of aleatory uncertainty. In contrast, the opposite trend in uncertainty rating was observed for tasks with random outcomes (e.g., stock price prediction, coin toss). Therefore, the experimental results provide support for Ülkümen and colleagues (2016)'s distinction of uncertainty and the general notion that there are different types of uncertainties (e.g., Kahneman & Tversky, 1982).

Chapters 3 and 4 investigated how advisees process the mismatch between external information ambiguity and adviser confidence using a basketball game outcome prediction task. More specifically, the advisers either expressed high confidence despite ambiguous external information or displayed uncertainty despite indicative external information. The objective of Chapter 3 was to examine how advisees flexibly adapt their confidence in response to unexpected adviser confidence and the role of expertise. Experiment 6 found that when advisers express uncertainty despite unambiguous external evidence (i.e., unambiguous trials), advisees decreased their confidence in spite of advice that agreed with their initial answer. The meta-analysis of the effect size found a small

but significant effect adviser, type of adviser (expert vs. novice), the difference between advisee's initial confidence and advisers' confidence (i.e., confidence distance) and the interaction of these two variables. In Experiments 7-8, it was expected that participants' change in confidence would be more difficult to detect due to the linguistic expressions of confidence, which was used either by the advisers or the advisees, and thus a larger number ( $N = 60$  vs.  $N = 30$  in Experiment 6) of participants were recruited. However, this may not have been sufficient considering the small effect size and future research could collect data from a larger sample to better detect the effects. Nevertheless, the findings of the experiments diverged from previous literature that generally finds that people become more confident in their opinion when others agree (Rader et al., 2017; Sah et al., 2013). Increase in confidence following agreeing advice is plausible according to a simple Bayesian model of decision-making as agreement is evidence that increases the probability of being correct (Bahrami et al., 2010). However, a more complex Bayesian model could suggest that advice is used to not only assess the accuracy of their immediate decision but also to update their prior belief about the task difficulty (Bate, 2007; Meyniel et al., 2015). In support of this notion, participants also updated their decision-making strategy, where participants became less reliant on two basketball statistics that were used less by the advisers. It is possible that participants deduced and adopted the adviser's weighting of the basketball statistics from adviser confidence. These findings suggest that people evaluate adviser confidence in relation to the external information. However, an alternative explanation was that participants perceive the task as an estimation task, where there is a specific "correct" confidence value. Therefore, confidence change could have been driven by a motivation to match their confidence to the adviser's confidence (i.e., confidence matching) rather than a reflection of genuine change in subjective confidence.

Hence, in two follow-up experiments, either the advisers (Experiment 7) or the participants (Experiment 8) used five linguistic expressions to convey their confidence. Linguistic expressions cover a range of confidence values and are open to interpretations (Hamm 1991; Harris & Corner, 2011). Therefore, participants are unlikely to perceive the task as an estimation task with correct point confidence value. In contrast to Experiment 6, in Experiment 7 and 8, participants generally tended to slightly increase their confidence following agreeing but uncertain advice when the external information was unambiguous. A possible explanation is that since linguistic expressions represent a range of confidence values, participants might have unnoticed when they were overconfident compared to the adviser and supposed that the advisers were expressing similar levels of confidence as themselves. Therefore, participants were more likely to increase their confidence following agreeing advice and only decrease their confidence when the advisers were considerably less confident. In addition, participants once again showed signs of learning from adviser confidence. Participants either numerically (Experiment 7) or significantly (Experiment 8) reduced their weighting of one basketball statistic that was less used by the advisers. Hence, experiments in this chapter demonstrated that advisees appraise adviser confidence in relation to the external information and flexibly use adviser confidence to adjust their immediate decision and general decision-making strategy. Additionally, in all experiments, participants were more sensitive to advice from an expert than a novice and perceived the expert as more competent. Thus, these findings consolidate the robust finding in advice taking that increased adviser expertise leads to greater advice adoption.

In Chapter 3, there were some hints of participants shifting their weighting of the external information (basketball statistics), which resembled the advisers' weighting. This pattern

was found despite the absence of performance feedback. Research in cue learning found that providing partial accuracy feedback is sufficient to instigate learning (De Klerk & Oppe, 1972; Hendrix & Dudycha, 2015; Stewart et al., 2012). Furthermore, providing probabilistic feedback facilitates learning of cue relationships (Little & Lewandowsky, 2009). It is possible that in Experiments 6-8, participants perceived adviser confidence as a proxy for accuracy feedback which resulted in learning the informativeness of the basketball statistics. Therefore, the role of adviser confidence in learning the validity of external information was further examined in Chapter 4. Once again, the task was to predict basketball game outcome given four basketball statistics. However, there were some methodological changes to facilitate learning, such as increased number of trials, only interacting with one adviser, and monetary incentives for accuracy (Experiment 10). Despite these methodological improvements, there were only small and non-significant changes in participants' use of the four basketball statistics. Although, in Experiment 10, participants' weighting of all four basketball statistics numerically decreased, which hints that advisees might adjust their decision-making strategy following advice only under certain circumstances. Furthermore, participants were able to somewhat estimate which external information was prioritised by the adviser and this estimation became more accurate after repeated interaction (Experiment 10). In addition, although most analysis in Experiments 6-10 found no difference in the use of the four basketball statistics, some experiments found small to medium effect sizes for offensive rebound (ORB), points scored the previous month, and percentage of wins this month. This suggests that there were small shifts in participants' use of these statistics in some cases. These puzzling and inconsistent findings could be further examined in future research using larger sample sizes. Overall, the results from Chapter 3 and 4 suggest that advisees may be able to learn from adviser confidence but only under specific conditions.

## **5.2 Implications**

### **5.2.1 Adviser Characteristics and Advice Adoption**

Numerous studies in advice-taking have found that greater adviser expertise and confidence leads to increased advice adoption. The findings in the current thesis also support this notion. Experiments 6-8 found that expert advisers were more influential than novice advisers. For example, participants changed their answer more often when an expert (vs. novice) disagreed. Additionally, the magnitude of confidence change was greater with an expert (Chapter 3). This is in line with various research that showed that advisees prefer expert advisers and are more likely to adopt their advice (Bailey et al., 2023; Bonnacio & Dalal, 2010; Sniezek et al., 2004; Jungermann & Fischer, 2014). Furthermore, post-advice satisfaction is also higher with experts than novices where people perceive an expert's advice as more valuable (Meshi et al., 2012). Similar tendencies were observed in Experiments 6-8. Perceived adviser competence, which was measured at the end of the experiment, was higher for the expert. This suggests that participants post-advice evaluation of advisers was more positive for the expert. This preference for expert advisers is a robust finding in Judgment and Decision-Making (JDM) research and is likely to replicate in future studies as well as other disciplines. For example, research in financial advice-taking showed that adviser's expertise influences perceived trustworthiness as well as future advice-seeking intention (Joiner et al., 2002). Preference for higher expertise also extends to lawyer and client (Elbers et al., 2012; Rosenthal, 1974) and adviser group and policymaker (Laage-Thomsen, 2022; Salacuse, 2018) relationships.

Furthermore, although measuring the effect of adviser confidence was not of main interest in the current thesis, the experiments showed that increased confidence led to greater

advice adoption. For instance, participants' tendency to decrease confidence when an adviser disagrees but increase confidence following agreeing advice was intensified based on the magnitude of the adviser's confidence. For instance, Experiments 6-8 showed that participants were attentive to adviser confidence, where participants often increased their confidence when an adviser was more confident than themselves but decreased their confidence when they were (highly) overconfident compared to the adviser. Furthermore, in addition to the analysis reported in Chapter 4, the correlation between adviser confidence and adviser influence (i.e., advice adoption) was positive ( $r = 0.42$ ) following agreement while the correlation was negative under disagreement ( $r = -0.23$ ). The findings support past research that has shown that advisees are more likely to adopt confidently given advice (Sniezek & Buckley, 1995; Sniezek & Van Swol 2001; Van Swol, 2009; Van Swol & Sniezek, 2001; Wang et al., 2025; Yaniv 1997). Furthermore, some studies have found that adviser confidence exerts greater influence on advisees than adviser accuracy or trust (Van Swol, 2005; Van Swol, 2009).

The current thesis affirms the importance of the magnitude of adviser confidence on advice adoption but also found hints that *how* adviser confidence is expressed could impact advice-taking. For example, some experiments in Chapter 2 found hints that participants preferred an adviser who used context-appropriate linguistic expressions to convey their confidence (e.g., Experiment 2-4, although Experiment 2 results were not replicated in Experiment 5). Some research in advice-taking also find that the linguistic expressions used in advice has an impact on advice adoption. More specifically, assertive expressions generally lead to greater advice adoption (Manian & Sheth, 2021), but this preference is context dependent. For example, Kronrod and colleagues (2012) found that assertive messages are more persuasive when communicating prominent environmental

issues, whereas suggestive languages are more effective for unfamiliar issues. This importance of context-appropriate adviser language (assertive vs. suggestive) has also been documented in medical advice-adoption (Kronrod et al., 2025). Lastly, the current thesis extends the research on acceptance of uncertain advice. Although confident advice is generally preferred, some researchers find that advice accompanied with some levels of uncertainty are also perceived positively (Du et al., 2024; Gaertig & Simmons, 2018; Loschelder et al., 2016; Schultze & Loschelder, 2020). This is because advice accompanied by some uncertainty (e.g., wider confidence interval) is considered more accurate and informative (Du et al., 2011). In Chapter 2, one adviser used words associated with uncertainty (i.e., probability, chance, likelihood) (Du et al., 2024; Gaertig & Simmons, 2018; Yates et al., 1996). However, participants were not averse to and, in some experiments, preferred this adviser. Comparison between certain and uncertain advice was not a main objective in the current thesis, yet the findings still contribute to the growing literature on uncertain advice and the results could be a blueprint for future research.

### **5.2.2 Processing advice in relation to external context**

The experimental chapters suggest that the advice-taking context influences advisees' processing of advice and the accompanying confidence. It was predicted that advisees would flexibly interpret adviser confidence and make inferences about the adviser's ability to assess the advice-taking context (Chapter 2) or their own decision-making (Chapter 3-4). The chapters together showed some evidence that the advice-taking contexts and adviser confidence are processed in conjunction. For example, participants showed some sensitivity the relationship between advisers' linguistic expression and contextual uncertainty (Chapter 2) or agreeing advice and external information ambiguity

(Chapter 3). Additionally, advisees may learn about the external information from adviser confidence under some conditions (Chapter 3-4).

Previous literature in advice-taking has demonstrated the importance of advice-taking context and external information. Researchers have observed significant change in advice-seeking and adoption depending on task difficulty. More specifically, increased task difficulty leads to greater advice-taking but people underweight advice as task difficulty decreases (Gino & Moore, 2006; Schrah et al., 2006). This positive correlation between advice-seeking and task difficulty is also found in other domains, such as medical decision-making (Kassirer et al., 2020) and firm management (Johnson et al., 2007; Mole et al., 2016). These findings demonstrate that advisees evaluate the advice-taking environment and adjust their advice adoption accordingly.

However, despite the clear impact of advice-taking context on degree of advice adoption, research in the area has been limited to the level of task difficulty. Furthermore, most research in the area has examined advice-taking in isolation where advisees have no or limited access to the objective information (Palmeira, 2020). Hence, the current thesis attempted to extend this literature by investigating the type of uncertainty in the advice-taking context (Chapter 2) and the interaction between external information and advice (Chapter 3-4). Only a small amount of research has examined the relationship between advice and the advice-taking context. For example, across seven studies, Palmeira (2020) investigated the perception of advisers who deviate from external information of quality (e.g., price, popularity). For example, wine critics ranked three different wines while either unaware or aware of the price. The wine critic's ranking was either in line with the price (i.e., rank most expensive as best) or deviated from the price information (e.g., rank

second most expensive as best). The studies found that when the wine critics were unaware of the price, the critic whose ranking matched the price information was perceived to have higher expertise than a critic whose ranking did not conflict with the price. On the other hand, when the price information was known, the ranking that varied from the price information was perceived more positively. This is because judgements that conflict with external information are viewed as less accurate but more independent and knowledgeable. Hence, advisees are less likely to seek further information and are more satisfied after receiving advice that deviates from the external information. However, advice that deviates extremely from the external information backfires (Palmeira, 2020). The findings demonstrate that people do not process adviser confidence in isolation but in relation to the external information. Some experiments in the current thesis also observed this tendency to process advice in relation to the advice-taking context. For example, perceived adviser competence and degree of advice adoption somewhat varied depending on the congruency between the advisers' linguistic expression and the type of contextual uncertainty (i.e., epistemic vs. aleatory) (e.g., Experiment 2-4, but not replicated in Experiment 5) or the level of aleatory uncertainty in the advice-taking context (Experiments 4-5). In addition, advisees did not always become more confident under agreeing advice, but change in confidence was dependent on the external information ambiguity (Experiments 6-8).

Research in marketing also finds that advisers who consider the advice-taking context are preferred and viewed as more competent. For example, people prefer advice that takes a more holistic approach by accounting for the relationship between different products. More specifically, advisers who give combinatory recommendations (e.g., bicycle and helmet) are perceived to have greater expertise than those who recommend single

products (D'Angelo & Valsesia, 2023; Zhang et al., 2024). The greater perception of expertise is not simply due to the increased number of advised products (i.e., recommended two products instead of one). More specifically, an adviser who demonstrated knowledge in equal number of products by suggesting substituting one product with another, was still perceived as less of an expert (D'Angelo & Valsesia, 2023). This higher perception of expertise was maintained even when the combinatory advice was less useful than a recommendation to substitute. Combinatory recommendations lead advisees to have a more positive attitude towards the product (Zhang et al., 2024) and become more confident in their decision, which subsequently increases purchase intention (Zhang et al., 2025). The preference for combinatory advice stems from the belief that these advisers have a greater depth (rather than breath) of knowledge and thus, advisees are also more likely to follow subsequent advice (D'Angelo & Valsesia, 2023). Advisees' inclination to evaluate adviser competence from advice was also found in some experiments in the current thesis. For instance, there were some signs of advisees inferring adviser competence from advisers' linguistic expression used in advice (Chapter 2). Furthermore, in Chapter 3, people were more likely to reduce their confidence when an expert (vs. a novice) agreed with their answer under ambiguous external information. This suggests that advisees process adviser competence and advice in conjunction. Indeed, Wang and colleagues (2025) found that advisees with greater expertise considered both the quality of advice and an adviser's expertise when assessing the adviser. It is possible that participants in Chapter 3 perceived unconfident advice from an expert as a sign of their misjudgement of the external information but believed that uncertain advice from a novice was due the novice's lack of competence. Although the experiments in the current thesis yielded mixed results, they have also demonstrated that the advice-taking context bears the potential to affect advice-adoption.

### 5.2.3 Inferences people make about adviser confidence

The current thesis found some evidence to suggest that advisees make inferences about adviser confidence. In Chapter 2, some experiments found small preferences for an adviser who uses context-appropriate linguistic expressions of confidence. This preference might be due to the assumption people make about the advisers' expertise. As previously mentioned, Price and Stone (2004) compared perception of an overly confident versus a reasonably calibrated financial adviser. The researchers found that people favour the extremely confident financial advisers, based on assumptions that confident advisers are more knowledgeable and make more correct judgements (see also Thomas & McFadyen, 1995). This phenomenon was termed confidence heuristic and has been consistently found in numerous studies (Phillips 1999; Pulford et al., 2018; Sniezek & Buckley, 1995; Sniezek & Van Swol 2001; Van Swol, 2009; Van Swol, 2011; Van Swol & Sniezek, 2001; Wang et al., 2025; Yaniv 1997). Although other studies demonstrate that people distrust advisers whose confidence and accuracy are not well calibrated (i.e., overconfident) (Sah et al., 2013; Stanciu & Fiser, 2022; Tenney et al., 2007; Tenney et al., 2008; Tenny & Spellman, 2011). The findings in Chapter 2 adds to the literature on confidence heuristic. In some experiments (Experiment 2, 4), participants perceived the adviser who used context appropriate linguistic expressions as more competent. Participants might have perceived the adviser who use context appropriate linguistic expression as more knowledgeable as their expressions demonstrate an understanding of the advice-taking context. This suggest that people make inferences about the adviser's competence even from *how* confidence is expressed. However, further research in the area is needed considering Experiment 5, which followed the same methodology as Experiment 2, failed to replicate the findings.

Researchers also suggest that people make other inferences from adviser confidence. For example, confidence is perceived as the degree of adviser's commitment to their answer (Vullioud et al., 2017). Additionally, advisees not only assume that confident advisers have higher expertise at the task level but also have greater general expertise (Sniezek & Van Swol, 2001). Advisees also infer the trustworthiness of an adviser from adviser confidence (Mayer et al., 1995). These findings show that people do not evaluate adviser confidence in isolation but make connections with other traits. Researchers have also demonstrated that people also appraise adviser confidence in relation to their own decision-making and make complex inferences. For instance, Pescetelli and Yeung (2021) found that advisee use the relationship between their own confidence, adviser's confidence and, whether the adviser agrees with their answer to deduce the adviser's accuracy and their calibration between accuracy and confidence. The results from Chapter 3 extends findings from previous literature and suggest that people use adviser's low confidence to evaluate one's decision-making. Participants not only used lower adviser confidence to re-evaluate their decision accuracy within a single trial but also seemed to infer that their overall task difficulty evaluation was erroneous. In particular, participants exhibited signs of learning external information validity from low adviser confidence. Although, the results from Chapter 4 suggest that learning may only occur under specific conditions. In sum, the current thesis provides another clue that advisees use adviser confidence to make diverse inferences about the adviser and themselves.

### **5.3 Limitations**

One limitation of the current thesis and the general judgement and decision-making research is that the experimental paradigm follows Judge-Adviser System (JAS) (Sniezek & Buckley, 1995). Use of JAS allows for careful experimental control and reduce

logistical issues (Bonnacio & Dalal, 2006; Bolger et al., 2023). Adopting JAS in the current thesis enable creation of precisely curated advice and to accurately measure advice adoption. However, this also raise questions about the generalisability of the results to real-world advice-taking. For example, one key limitation is that the communication between the adviser and advisee was one-sided, where only the adviser gave their opinion, and advisees did not have a chance to interact with the adviser. This type of interaction may resemble some advice-taking situations such as, reading online reviews but is far from advice-taking contexts where the advisee and the adviser can freely communicate. In general, advisees would be able to communicate (at least to some degree) with the adviser in the real-world, which may impact the quality of advice. For example, Nückles and colleagues (2005) found that when advisees informed the adviser about their level of knowledge, experts gave a more tailored and useful advice. Additionally, studies find that posts on online-forums with greater self-disclosure result in greater empathetic behaviour from other users and increase the information received (Lee et al., 2023; Lei et al., 2023; Pan et al., 2018; Zou et al., 2024). These findings demonstrate that advisers are responsive to the information divulged by the advisee and may change the content of their advice accordingly. For example, during basketball game outcome prediction (Chapter 3-4), the adviser might adjust the amount of reasoning they provide based on the advisee's pre-existing knowledge. In the current experiments, the focus was the effect of adviser expertise and confidence thus the content of the advice was minimal. However, future research could examine the relationship between dynamic communication and evolution of advice content. For example, the perceived level of epistemic and aleatory uncertainty in the advice-taking context could be subjective and an adviser could ask for the advisee's perception and tailor their advice accordingly.

Furthermore, all experiments in the current thesis were conducted online via Prolific (<https://www.prolific.com/>). Most research suggest that online and lab-based experiments yield similar results (Arechar et al., 2018; Buso et al., 2021; Reips 2001) and effect sizes (Sauter et al., 2022). This has been found across various tasks, such as False Belief tasks (Schidelko et al., 2021), mental comparison (Ollesch et al., 2006), Dictator Game, and cognitive refection (Prissé & Jorrat, 2022). However, online participants tend to pay less attention to tasks (Albert & Smilek, 2023; [Clifford](#) & Jerit, 2014; Hergueux & Jacquemet, 2015) and complete tasks faster (Hergueux & Jacquemet, 2015). Furthermore, online participants have greater difficulty understanding task instructions as task complexity increases (Finley & Penningroth, 2015). For example, online participants were less accurate in a problem-solving task compared to in-lab participants (Dandurand et al., 2008). These limitations of online experiments are likely to only have a small impact in the current experiments as participants on Prolific are generally attentive compared to other online samples (Albert & Smilek, 2023; [Uittenhove](#) et al., 2023) and the tasks were simple. On the other hand, although the simple task structure is advantageous in measuring the effects of the manipulated variable (e.g., confidence, expertise), field experiments may be better suited to examine complex advice-taking dynamics and examine various factors such as advice solicitation (Kämmer et al., 2023). The focus of the thesis was to examine advice adoption, and thus the current research paradigms were suitable. However, future research combining different research methods could provide a fuller picture of advice-taking.

A related limitation of the current experiments is that participants were provided with advice and advice solicitation was not investigated. Previous research suggests that people are less likely to seek advice when they are highly confident (Carlebach & Yeung,

2023; Kramer, 2016; Pescetelli et al., 2021; Snijders et al., 2022; Soll & Larrick, 2009) and when decision stakes are low (Løhre & Halkjelsvik, 2024). Therefore, participants might have seldom sought advice in some tasks or trials when given the choice. For example, in the coin toss (aleatory) task in Experiments 4 and in Experiments 6-8 where external information was unambiguous, participants were generally confident and advice seeking might have been low. Indeed, adviser influence in the coin toss (aleatory) task was much lower compared to other tasks in Chapter 2. On the other hand, people are more likely to seek external input for knowledge-based tasks (e.g., political knowledge questions) (Clifford & Jerit, 2014). Therefore, advice solicitation might be high in epistemic tasks (e.g., historic dates) or for trials in Experiments 6-10 where the external information was ambiguous and subsequent participants' confidence was low. In addition, there are large individual variations in advice-seeking (Pescetelli et al., 2021). This tendency may also be driven by internal factors (e.g., self-efficacy, shyness, mood), similar to factors that influence advice-adoption (as discussed in Chapter 1). In support of this, researchers have found that emotions, such as fear, (Ferrer et al., 2021) and clarity of one's self-concept (Duan et al., 2020) influence advice-seeking. Although advice solicitation was not investigated in the current empirical chapters, it is undeniable that it is a significant aspect of advice-taking. Future research in JDM could further examine how individual differences impact advice-seeking.

Another possible difference between the current experiments and real-life advice-taking is that advisees were aware that advice adoption was private. This was to accurately measure advisees' internal change in perception or belief. However, advice-taking could occur in public settings and advisees may be susceptible to pressure to conform (e.g., due to social norm of reciprocal advice-taking) and comply (e.g., pressure from adviser to

accept advice). Seminal work on social conformity by Asch (1951) suggest that peoples' private and public belief and behaviour could vary. Furthermore, people sacrifice accuracy in the presence of social pressure (Asch, 1951; Asch, 1956). In addition, advisers are generally sensitive to whether their advice was accepted and how helpful their advice was (Li et al., 2024). For instance, advisers react negatively to rejection of their advice, such as perceiving the advisee as less competent (Palmeira & Romero Lopez, 2023) or expressing lower intention to give advice in the future (Belkin & Kong, 2018; Blunden et al., 2019). These behaviour by an adviser would put implicit pressure on advisees to comply (Cialdini & Goldstein, 2004). Indeed, Sciandra (2024) found that advisees increase their advice adoption when it is made public. A pair of participants recommended to each other in which combination they should select five pieces of chocolate from two types (e.g., 1 Kit Kat and 4 Hershey bars; 3 Kit Kats and 2 Hershey bars). After, the pair were given a bowl of chocolates, and one member was asked to pick five pieces in the presence of their partner. On the other hand, the second member selected chocolates in private. The study found that the first member was significantly more likely to adopt their partners recommendation. This increased advice adoption was found in other contexts (e.g., historic dates, purchasing a blender) and also when post-advice decision was viewed by other advisees instead of the advisers. Greater advice adoption in public (vs. private) setting is likely to be driven by one's desire to maintain a positive social relationship and self-image (Cialdini & Goldstein, 2004; Sciandra, 2024). In the context of the current experiments, advice adoption could be greater under public advice-taking. For instance, in Experiments 7-8, participants only decreased their confidence when they were considerably overconfident compared to the adviser. However, change in participants confidence would have been greater if advisers were to view advisees' post-advice confidence. Additionally, participants might have decreased their advice more frequently

and even after novice's advice. It is worth noting that increasing advice adoption would be beneficial in the current experiments as the advisers were generally accurate. However, this may not always be the case in real-life. For example, some financial advisers could help make more rational decisions and help increase one's asset (Montmarquette & Viennot-Briot, 2015; Sunder et al., 2024), but some advisers perform no better than laymen (Linnainmaa et al., 2021) and could be deceptive (Egan et al., 2019). Public advice-taking could prompt advisees to blindly adhere to incompetent and malicious advisers which would decrease their decision accuracy.

Furthermore, the adviser and advisee had no prior relationship, but this would not always be the case. For instance, in real-life, an adviser and an advisee may have prior relationships (e.g., friends, colleagues) and are likely to interact again in the future (Kämmer et al., 2023; Rader et al., 2017). In such cases, advice-giving tends to be a bidirectional process and the adviser may have access to whether their advice was implemented. As previously mentioned, research shows there is an element of reciprocity with advice-taking, where people are compelled to adopt advice from a person who accepted their advice (Mohamodi et al., 2018). Furthermore, there are higher stakes for ignoring advice when there are existing social relationships as advisers are more sensitive to advice rejected by friends compared to strangers (Zhang et al., 2022) and rejecting advice could deteriorate the whole relationship (Rader et al., 2017). Additionally, in organisations, other factors such as status and reputation may play a role (Kämmer et al., 2023). For example, an adviser might be a superior and thus, have greater authority and power (Kämmer et al., 2023), which amplifies social compliance (Cialdini & Goldstein, 2004). Advisers are also sensitive to the advisee's status and are more disappointed when a high (vs. low) status individual dismisses their advice (Li et al., 2024). However,

powerful individuals are reluctant to accept advice as they perceive advice as a threat to their autonomy and reputation (Liang & Xue, 2022; Paik, 2020; Tost et al., 2012; De Wit et al., 2017, See et al., 2011). The results of the current experiments might have varied if advice-taking occurred between pairs with existing relationship. The implicit social pressure to comply and greater adviser's authority would have led participants to pay greater attention to advice. This in turn might have led participants to better detect the external information prioritised by the adviser in Chapter 4. Furthermore, participants would have been more likely to adopt the adviser's use of the external information, and a greater learning effect could have been detected. On the other hand, the findings in Chapter 3 (i.e., decrease in confidence despite agreement under ambiguous external information and learning from adviser confidence) might have been weaker if advisees had higher status than the adviser.

Lastly, the participants in the current experiments were recruited from the United Kingdom (U.K.) or the United States (U.S.) whose primary language was English. This was to ensure that all participants could fully understand the task (e.g., nuance of linguistic expressions in Experiments 1-5) and prevent cultural factors influencing advice adoption. However, the U.K. and U.S. has generally been categorised as individualistic cultures (Sun et al., 2004) and researchers suggest that the individualism-collectivism cultural dimension may influence advice-taking and advice-giving. For example, Schultze and Chen (2025) found that Chinese (collectivist culture) people were more likely to incorporate advice than Germans (individualists). Furthermore, advisees from collectivist cultures (Chinese) placed greater importance on adviser characteristic during advice-taking whereas people from individualist cultures (Americans) focused on advice content. Similarly, individualism and collectivism dimension also influence motivation

for advice-seeking, where European Canadians are driven to gain information or to solve problems whereas Chinese advisees place greater importance on enhancing relationships with advisers (Ji et al., 2017). This difference in behaviour may be driven by the different perception of advice where advice is considered as a friendly and helpful act in collectivist cultures but often perceived as an intrusion in individualist cultures (Hosni, 2020). On the other hand, people from collectivist cultures still practice egocentric discounting (i.e., place greater weight on own opinion) (Schultze & Chen, 2025) and some studies found no difference in degree of egocentric discounting by culture (Mercier et al., 2012). Moreover, Feng and colleagues (2016) suggest that individuals' perception of advice (e.g., helping the distressed person regain confidence) is a stronger predictor of advice-giving behaviour than culture. Therefore, findings from previous literature suggest that the general pattern of the current experimental results are likely to be similar regardless of cultural background of the participants. However, participants from collectivist cultures might have been more sensitive to advice and adviser status. For example, a greater effect of adviser expertise might have been observed in Experiments 6-8.

#### **5.4 Future directions**

One possible future research direction is the relationship between adviser confidence and the advice-taking context. Research in judgement and decision-making (JDM) have largely been focused on understanding the degree of advice adoption in isolation (Bonnacio & Dalal, 2006). However, some past research has found contextual factors such as task difficulty (Schrah et al., 2006), whether advice was solicited or not (Paik, 2020; Van Swol) and consistency with external cues of quality (Palmeira, 2020) to play a role. The current thesis found some evidence to suggest that advisees are sensitive to the relationship between contextual uncertainty and advisers' linguistic expression

(Chapter 2) or under which condition (external information is ambiguous vs. unambiguous) the adviser agrees with their answer (Chapter 3). Future studies could further explore the impact of contextual factors on other advice-taking behaviours. For example, it was found that increase in aleatory uncertainty leads to reduced advice-adoption. Researchers could examine if this reduced advice-taking applies to all types of advice-seeking (e.g., exploratory vs. confirmatory).

Further investigation into the advice-taking context would also help better understand under which conditions certain advice-taking behaviours are elicited. For instance, Chapter 3 found hints of learning the validity of external information from advisers' confidence, but this learning was weakened following methodological changes in Chapter 4. This suggests that learning from adviser confidence may only occur under specific conditions. For example, advisees may only process adviser confidence deliberately when they need to evaluate an adviser's competence (as did in Chapter 3 to compare novice and expert advisers). Other factors such as the information provided (e.g., amount, quality) or the magnitude of discrepancy between external information ambiguity and advice. Further research would not only help clarify when learning from adviser confidence occurs but better identify the role of advice-taking context. Some past research findings suggest that advice-taking contexts could lead to different outcomes despite identical advice content. For example, Pescetelli and Yeung (2021) showed that advisees could learn an adviser's confidence and accuracy calibration in the absence of external feedback, but this ability was reduced when the advisers and advisees shared the same judgement biases (e.g., same information source). Additionally, investors are more trusting of robo-financial advisers due to low risk of fraud or conflict of interest. However, this trust decreases when the market is volatile due to the difficulty accessing the robo-adviser's

decision-making process (Tan, 2020). This demonstrates that advice evaluation and perception is influenced by the advice-taking context. Research that varies the advice-taking context could help better pinpoint which factors elicit different strategies of advice processing. This could also help advisers to effectively communicate information in relation to the advice-taking context.

Another potential line of research is advice-taking between pairs with pre-existing social relationships under dynamic communication. This was a limitation in the current thesis and more widely, the general advice-taking literature (Bonnacio & Dalal, 2006; Kämmer et al., 2023). As previously discussed, in real-life advice-taking, advisees and advisers can generally interact freely and prior social relationships may impact advice-taking. For instance, advice adoption in the current experiments might have been greater if social factors played a role, which could have facilitated detection of the experimental effects. For example, the preference for context matching linguistic expressions of confidence might have been more apparent (Chapter 1). Another difference might be that greater communication between adviser and advisee could reduce the need for advisees to deduce the adviser's decision-making strategy (Chapter 3) and the experimental focus would shift to whether advisees are willing to fully adopt an adviser's decision-making strategy or only partially adopt. In the current thesis, understanding the informational aspect of advice was of interest rather than the social aspect. However, it is undeniable that the social aspect plays a critical role in advice-taking. Therefore, future studies investigating the interaction between the informational and social features of advice would lead to deeper understanding of advice-taking in real-life. For example, a rising research method in judgement and decision-making is comparing advisee's reaction to advice from humans versus Artificial Intelligence (AI) (e.g., Jin et al., 2025; Osborne & Bailey, 2025;

Vodrahalli et al., 2022). Assuming the content of advice is identical, the difference in advice-adoption between the two agents could be attributed to the social aspects (e.g., reciprocity) of advice-taking. Extending this research method, chatbots could also be used to research dynamic interactions. Chatbots may be perceived similar to an expert in terms of competence but how advisees respond may deviate from traditional advice-taking.

## **5.5 Conclusion**

The current thesis aimed to better understand how adviser confidence is processed in relation to the advice-taking context and possible inferences advisees make about adviser confidence. The experimental findings suggest that the advice-taking context influences general advice-taking behaviour (e.g., degree of advice-adoption, decrease in confidence despite agreement). Furthermore, adviser confidence may be used to infer adviser characteristic (e.g., competence) and features of the advice-taking context (e.g., task difficulty, validity of external information). However, the findings were inconsistent suggesting that these behaviours may only be elicited under specific conditions. Nonetheless, the experimental results demonstrate that there are multiple facets of adviser's confidence and an adviser's confidence is not processed in isolation. Future research could further investigate the specific features of the advice-taking context that are linked to adviser confidence to better understand when and how complex processing of adviser confidence is evoked. Research in the area would not only enrich the judgement and decision-making (JDM) literature but would benefit all disciplines related to advice-taking.

# References

---

- Ache, F., Rader, C., & Hütter, M. (2020). Advisors want their advice to be used – but not too much: An interpersonal perspective on advice taking. *Journal of Experimental Social Psychology, 89*, 103979. <https://doi.org/10.1016/j.jesp.2020.103979>
- Aguilar-Lleyda, D., Lemarchand, M., & De Gardelle, V. (2020). Confidence as a priority signal. *Psychological Science, 31*(9), 1084–1096. <https://doi.org/10.1177/09567976209250>
- Ajzen, I., Fishbein, M., Lohmann, S., & Albarracín, D. (2018). The influence of attitudes on behavior. *The handbook of attitudes, volume 1: Basic principles*, 197–255.
- Albert, D. A., & Smilek, D. (2023). Comparing attentional disengagement between Prolific and MTurk samples. *Scientific Reports, 13*(1), 20574. <https://doi.org/10.1038/s41598-023-46048-5>
- Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. In M. D. Alicke, D. A. Dunning, & J. I. Krueger (Eds.), *The self in social judgment* (pp. 85–106). Psychology Press.
- Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology, 68*(5), 804–825. <https://doi.org/10.1037/0022-3514.68.5.804>
- Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics, 21*(1), 99–131. <https://doi.org/10.1007/s10683-017-9527-2>
- Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In *Groups, leadership and men; research in human relations* (pp. 177–190). Carnegie Press.
- Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70*(9), 1–70. <https://doi.org/10.1037/h0093718>
- Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally Interacting Minds. *Science, 329*(5995), 1081–1085. <https://doi.org/10.1126/science.1185718>
- Bailey, P. E., Leon, T., Ebner, N. C., Moustafa, A. A., & Weidemann, G. (2023). A meta-analysis of the weight of advice in decision-making. *Current Psychology, 42*(28), 24516–24541. <https://doi.org/10.1007/s12144-022-03573-2>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bate, A. (2007). Bayesian confidence propagation neural network. *Drug safety, 30*, 623–625. <https://doi.org/10.2165/00002018-200730070-00011>
- Belkin, L. Y., & Kong, D. T. (2018). Implications of advice rejection in repeated exchanges: Advisor responses and advisee gratitude expression as a buffer. *Journal of Experimental Social Psychology, 78*, 181–194. <https://doi.org/10.1016/j.jesp.2018.03.012>
- Benjamin, D., & Budescu, D. V. (2015). Advice from Experience: Communicating Incomplete Information Incompletely. *Journal of Behavioral Decision Making, 28*(1), 36–49. <https://doi.org/10.1002/bdm.1825>

- Billings, R. S., & Scherer, L. L. (1988). The effects of response mode and importance on decision-making strategies: Judgment versus choice. *Organizational Behavior and Human Decision Processes*, 41(1), 1-19. [https://doi.org/10.1016/0749-5978\(88\)90043-X](https://doi.org/10.1016/0749-5978(88)90043-X)
- Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, 37(1), 48–74. <https://doi.org/10.1037/0022-3514.37.1.48>
- Blunden, H., Logg, J. M., Brooks, A. W., John, L. K., & Gino, F. (2019). Seeker beware: The interpersonal costs of ignoring advice. *Organizational Behavior and Human Decision Processes*, 150, 83-100. <https://doi.org/10.1016/j.obhdp.2018.12.002>
- Bobokhidze, E., Weech, M., Fallaize, R., Hwang, F., & Lovegrove, J. (2024). Subjective feedback from the EatWellUK-2 study: Insights for personalised nutrition apps. *Proceedings of the Nutrition Society*, 83(OCE4), E440. <https://doi.org/10.1017/S0029665124006785>
- Bolger, F., Rowe, G., Hamlin, I., Belton, I., Crawford, M., Sissons, A., ... & Wright, G. (2023). Virtuous opinion change in structured groups. *Judgment and Decision Making*, 18, e25.
- Bonaccio, S., & Dalal, R. S. (2006). Advice-taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational behavior and human decision processes*, 101(2), 127-151. <https://doi.org/10.1016/j.obhdp.2006.07.001>
- Bonaccio, S., & Dalal, R. S. (2010). Evaluating advisors: A policy-capturing study under conditions of complete and missing information. *Journal of Behavioral Decision Making*, 23(3), 227-249. <https://doi.org/10.1002/bdm.649>
- Borland, R., Balmford, J., & Hunt, D. (2004). The effectiveness of personally tailored computer-generated advice letters for smoking cessation. *Addiction*, 99(3), 369-377. <https://doi.org/10.1111/j.1360-0443.2003.00623.x>
- Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, 12(1), 11–30. <https://doi.org/10.1037/1076-898X.12.1.11>
- Briazu, R. A., Bell, L., Dodd, G. F., Blackburn, S., Massri, C., Chang, B., ... & McCloy, R. (2024). The effectiveness of personalised food choice advice tailored to an individual's socio-demographic, cognitive characteristics, and sensory preferences. *Appetite*, 201, 107600.
- Brooks, A. W., Gino, F., & Schweitzer, M. E. (2015). Smart people ask for (my) advice: Seeking advice boosts perceptions of competence. *Management Science*, 61(6), 1421-1435. <https://doi.org/10.1287/mnsc.2014.2054>
- Budescu, D. V., & Rantilla, A. K. (2000). Confidence in aggregation of expert opinions. *Acta Psychologica*, 104(3), 371–398. [https://doi.org/10.1016/S0001-6918\(00\)00037-8](https://doi.org/10.1016/S0001-6918(00)00037-8)
- Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. In *Psychology of learning and motivation* (Vol. 32, pp. 275-318). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60313-8](https://doi.org/10.1016/S0079-7421(08)60313-8)
- Buso, I. M., Cagno, D. D., Ferrari, L., Larocca, V., Lorè, L., Marazzi, F., Panaccione, L., & Spadoni, L. (2021). Lab-like findings from online experiments. *Journal of the Economic Science Association*, 7(2), 184–193. <https://doi.org/10.1007/s40881-021-00114-8>

- Camerer, C., & Lovallo, D. (1999). Overconfidence and Excess Entry: An Experimental Approach. *American Economic Review*, 89(1), 306–318. <https://doi.org/10.1257/aer.89.1.306>
- Capuano, C., & Chekroun, P. (2024). A Systematic Review of Research on Conformity. *International Review of Social Psychology*, 37(1). <https://doi.org/10.5334/irsp.874>
- Carlebach, N., & Yeung, N. (2023). Flexible use of confidence to guide advice requests. *Cognition*, 230, 105264. <https://doi.org/10.1016/j.cognition.2022.105264>
- Çelen, B., Kariv, S., & Schotter, A. (2010). An Experimental Test of Advice and Social Learning. *Management Science*, 56(10), 1687–1701. <https://doi.org/10.1287/mnsc.1100.1228>
- Chan, I. C. C., Lam, L. W., Chow, C. W., Fong, L. H. N., & Law, R. (2017). The effect of online reviews on hotel booking intention: The role of reader-reviewer similarity. *International Journal of Hospitality Management*, 66, 54-65. <https://doi.org/10.1016/j.ijhm.2017.06.007>
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55(1), 591-621. <https://doi.org/10.1146/annurev.psych.55.090902.142015>
- Clayson, D. E. (2005). Performance overconfidence: metacognitive effects or misplaced student expectations?. *Journal of Marketing Education*, 27(2), 122-129. <https://doi.org/10.1177/0273475304273525>
- Clifford, S., & Jerit, J. (2014). Is There a Cost to Convenience? An Experimental Comparison of Data Quality in Laboratory and Online Studies. *Journal of Experimental Political Science*, 1(2), 120–131. <https://doi.org/10.1017/xps.2014.5>
- Cohen, M., Etner, J., & Jeleva, M. (2008). Dynamic decision making when risk perception depends on past experience. *Theory and Decision*, 64, 173-192. <https://doi.org/10.1007/s11238-007-9061-3>
- Collins, H. (2016). Three dimensions of expertise. *Expertise, communication, and organizing*, 12, 253-273. <https://doi.org/10.1007/s11097-011-9203-5>
- Cooper, D. A. (2024). Diffusion of Responsibility for Actions With Advice. *Journal of Behavioral Decision Making*, 37(4), e2415. <https://doi.org/10.1002/bdm.2415>
- Cottam, E., & Jha, P. P. (2025). Advising elite performers: the role of intuition, trust and expertise. *Management Decision*, 63(3), 1018-1036. <https://doi.org/10.1108/MD-11-2023-2076>
- D'Angelo, J. K., & Valsesia, F. (2023). You should try these together: Combinatory recommendations signal expertise and improve product attitudes. *Journal of Marketing Research*, 60(1), 155-169. <https://doi.org/10.1177/00222437221111344>
- Dalal, R. S., & Bonaccio, S. (2010). What types of advice do decision-makers prefer? *Organizational Behavior and Human Decision Processes*, 112(1), 11–23. <https://doi.org/10.1016/j.obhdp.2009.11.007>
- Dandurand, F., Shultz, T. R., & Onishi, K. H. (2008). Comparing online and lab methods in a problem-solving experiment. *Behavior Research Methods*, 40(2), 428–434. <https://doi.org/10.3758/BRM.40.2.428>
- Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81(2), 95–106. <https://doi.org/10.1037/h0037613>
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical Versus Actuarial Judgment. *Science*, 243(4899), 1668–1674. <https://doi.org/10.1126/science.2648573>

- De Hooge, I. E., Verlegh, P. W., & Tzioti, S. C. (2014). Emotions in advice taking: The roles of agency and valence. *Journal of Behavioral Decision Making*, 27(3), 246-258. <https://doi.org/10.1002/bdm.1801>
- De Klerk, L. F. W., & Oppe, S. (1972). Functional cue learning under different probabilistic feedback conditions. *Psychonomic Science*, 29(6), 377-380. <https://doi.org/10.3758/BF03336606>
- De Wit, F. R., Scheepers, D., Ellemers, N., Sassenberg, K., & Scholl, A. (2017). Whether power holders construe their power as responsibility or opportunity influences their tendency to take advice from others. *Journal of Organizational Behavior*, 38(7), 923-949. <https://doi.org/10.1002/job.2171>
- Derex, M., Feron, R., Godelle, B., & Raymond, M. (2015). Social learning and the replication process: an experimental investigation. *Proceedings of the Royal Society B: Biological Sciences*, 282(1808), 20150719. <https://doi.org/10.1098/rspb.2015.0719>
- Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological science*, 29(5), 761-778. <https://doi.org/10.1177/0956797617744771>
- Desender, K., Donner, T. H., & Verguts, T. (2021). Dynamic expressions of confidence within an evidence accumulation framework. *Cognition*, 207, 104522. <https://doi.org/10.1016/j.cognition.2020.104522>
- Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, 51(3), 629-636. <https://doi.org/10.1037/h0046408>
- Dove, G., Balestra, M., Mann, D., & Nov, O. (2020). Good for the Many or Best for the Few? A Dilemma in the Design of Algorithmic Advice. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2), 1-22.
- Droll, J. A., Abbey, C. K., & Eckstein, M. P. (2009). Learning cue validity through performance feedback. *Journal of Vision*, 9(2):18, 1-22, <https://doi.org/10.1167/9.2.18>
- Du, N., Budescu, D. V., Shelly, M. K., & Omer, T. C. (2011). The appeal of vague financial forecasts. *Organizational Behavior and Human Decision Processes*, 114(2), 179-189. <https://doi.org/10.1016/j.obhdp.2010.10.005>
- Du, X., He, R., Wang, Y., & Wang, J. (2024). The impact of advice uncertainty and individual regulatory modes on advice taking. *European Journal of Social Psychology*, 54(4), 933-945. <https://doi.org/10.1002/ejsp.3063>
- Duan, J., Xu, Y., & Van Swol, L. M. (2021). Influence of self-concept clarity on advice seeking and utilisation. *Asian Journal of Social Psychology*, 24(4), 435-444. <https://doi.org/10.1111/ajsp.12435>
- Dukas, R. (2019). Animal expertise: mechanisms, ecology and evolution. *Animal behaviour*, 147, 199-210.
- Duncan, K. D., & Shohamy, D. (2016). Memory states influence value-based decisions. *Journal of Experimental Psychology: General*, 145(11), 1420.
- Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process?. *Educational research review*, 1(1), 3-14. <https://doi.org/10.1016/j.edurev.2005.11.001>
- Efklides, A. (2014). How does metacognition contribute to the regulation of learning? An integrative approach. *Psihologijske teme*, 23(1), 1-30.
- Egan, M., Matvos, G., & Seru, A. (2019). The market for financial adviser misconduct. *Journal of Political Economy*, 127(1), 233-295.

- Elbers, N. A., van Wees, K. A. P. C., Akkermans, A. J., Cuijpers, P., & Bruinvels, D. J. (2012). Exploring Lawyer–Client Interaction. *Psychological Injury and Law*, 5(1), 89–94. <https://doi.org/10.1007/s12207-012-9120-0>
- Evans, A. M., Stavrova, O., & Rosenbusch, H. (2021). Expressions of doubt and trust in online user reviews. *Computers in Human Behavior*, 114, 106556. <https://doi.org/10.1016/j.chb.2020.106556>
- Feng, B., & MacGeorge, E. L. (2010). The influences of message and source factors on advice outcomes. *Communication Research*, 37(4), 553–575. <https://doi.org/10.1177/0093650210368258>
- Ferrer, R. A., Ellis, E. M., Orehek, E., & Klein, W. M. P. (2022). Fear increases likelihood of seeking decisional support from others when making decisions involving ambiguity. *Journal of Behavioral Decision Making*, 35(3), e2266. <https://doi.org/10.1002/bdm.2266>
- Finley, A. J., & Penningroth, S. L. (2015). Online versus in-lab: pros and cons of an online prospective memory experiment. In A. M. Columbus (Ed.), *Advances in Psychology Research*, vol. 113 (pp. 135-162). Hauppauge, NY: Nova Science Publishers, Inc.
- Fischer, I., & Harvey, N. (1999). Combining forecasts: What information do judges need to outperform the simple average? *International Journal of Forecasting*, 15(3), 227–246. [https://doi.org/10.1016/S0169-2070\(98\)00073-9](https://doi.org/10.1016/S0169-2070(98)00073-9)
- Fisher, H. F., & Tally, J. (1997). Isoergonic cooperativity in glutamate dehydrogenase complexes: a new form of allostery. *Biochemistry*, 36(36), 10807-10810.
- Fleming, S. M. (2024). Metacognition and Confidence: A Review and Synthesis. *Annual Review of Psychology*, 75(1), 241–268. <https://doi.org/10.1146/annurev-psych-022423-032425>
- Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review*, 124(1), 91–114. <https://doi.org/10.1037/rev0000045>
- Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, 4(2), 195–202. <https://doi.org/10.1037/h0023552>
- Friedemann, M., Bang, D., & Yeung, N. (2025). Normative and informational confidence matching. *Journal of Experimental Psychology: General*, 154(3), 759–774. <https://doi.org/10.1037/xge0001706>
- Gaertig, C., & Simmons, J. P. (2018). Do people inherently dislike uncertain advice?. *Psychological Science*, 29(4), 504-520. <https://doi.org/10.1177/0956797617739369>
- Gaertig, C., & Simmons, J. P. (2023). Are people more or less likely to follow advice that is accompanied by a confidence interval?. *Journal of Experimental Psychology: General*, 152(7), 2008-2025
- GALTON, F. Vox Populi . *Nature* 75, 450–451 (1907). <https://doi.org/10.1038/075450a0>
- Gardner, P. H., & Berry, D. C. (1995). The effect of different forms of advice on the control of a simulated complex system. *Applied Cognitive Psychology*, 9(7), S55–S79. <https://doi.org/10.1002/acp.2350090706>
- Gaube, S., Suresh, H., Raue, M., Lermer, E., Koch, T. K., Hudecek, M. F., ... & Colak, E. (2023). Non-task expert physicians benefit from correct explainable AI advice when reviewing X-rays. *Scientific reports*, 13(1), 1383.
- Gentile, M., Linciano, N., & Soccorso, P. (2016). *Financial advice seeking, financial knowledge and overconfidence: Evidence from the Italian market* (CONSOB

- Working Paper No. 83). Commissione Nazionale per le Società e la Borsa. <http://dx.doi.org/10.2139/ssrn.2802589>
- Gibbons, A. M., Sniezek, J. A., & Dalal, R. S. (2003, November). Antecedents and consequences of unsolicited versus explicitly solicited advice. In D. Budescu (Chair), *Symposium in Honor of Janet Sniezek*. Symposium presented at the annual meeting of the Society for Judgment and Decision Making, Vancouver, BC.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, 62(1), 451-482. <https://doi.org/10.1146/annurev-psych-120709-145346>
- Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making*, 20(1), 21–35. <https://doi.org/10.1002/bdm.539>
- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.*, 30(1), 535-574. <https://doi.org/10.1146/annurev.neuro.29.051605.113038>
- Goldsmith, D. J., & Fitch, K. (1997). The normative context of advice as social support. *Human communication research*, 23(4), 454-476. <https://doi.org/10.1111/j.1468-2958.1997.tb00406.x>
- Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. *Elife*, 5, e13388. <https://doi.org/10.7554/eLife.13388>
- Gustafson, A., & Rice, R. E. (2019). The effects of uncertainty frames in three science communication topics. *Science Communication*, 41(6), 679-706.
- Hamm, R. M. (1991). Selection of verbal probabilities: A solution for some problems of verbal probability expression. *Organizational Behavior and Human Decision Processes*, 48(2), 193-223. [https://doi.org/10.1016/0749-5978\(91\)90012-I](https://doi.org/10.1016/0749-5978(91)90012-I)
- Haran, U., Mazar, A., Hurwitz, M., & Moran, S. (2022). Confidently at your service: Advisors alter their stated confidence to be helpful. *Organizational Behavior and Human Decision Processes*, 171, 104154. <https://doi.org/10.1016/j.obhdp.2022.104154>
- Harris, A. J., & Corner, A. (2011). Communicating environmental risks: Clarifying the severity effect in interpretations of verbal probability expressions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(6), 1571-1578.
- Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, 70(2), 117–133. <https://doi.org/10.1006/obhd.1997.2697>
- Harvey, N., & Harries, C. (2004). Effects of judges' forecasting on their later combination of forecasts for the same outcomes. *International Journal of Forecasting*, 20(3), 391–409. <https://doi.org/10.1016/j.ijforecast.2003.09.012>
- Harvey, N., Harries, C., & Fischer, I. (2000). Using Advice and Assessing Its Quality. *Organizational Behavior and Human Decision Processes*, 81(2), 252–273. <https://doi.org/10.1006/obhd.1999.2874>
- Hendrix, W. H., & Dudycha, A. L. (1981). Feedforward and Feedback in Multiple Cue Probability Learning-Facilitating or Debilitating?. *The Journal of Experimental Education*, 49(3), 137-146. <https://doi.org/10.1080/00220973.1981.11011777>
- Hergueux, J., & Jacquemet, N. (2015). Social preferences in the online laboratory: A randomized experiment. *Experimental Economics*, 18(2), 251–283. <https://doi.org/10.1007/s10683-014-9400-5>
- Herling, R. W. (2000). Operational definitions of expertise and competence. *Advances in developing human resources*, 2(1), 8-21.

- Hertz, U., Bell, V., & Raihani, N. (2021). Trusting and learning from others: immediate and long-term effects of learning from observation and advice. *Proceedings of the Royal Society B*, 288(1961), 20211414.
- Hertz, U., Palminteri, S., Brunetti, S., Olesen, C., Frith, C. D., & Bahrami, B. (2017). Neural computations underpinning the strategic management of influence in advice giving. *Nature Communications*, 8(1), 2191
- Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision research*, 37(15), 2133-2141. [https://doi.org/10.1016/S0042-6989\(97\)00043-6](https://doi.org/10.1016/S0042-6989(97)00043-6)
- Horstmann, G. (2015). The surprise–attention link: A review. *Annals of the New York Academy of Sciences*, 1339(1), 106-115. <https://doi.org/10.1111/nyas.12679>
- Hosni, H. R. (2020). Advice giving in Egyptian Arabic and American English: A cross-linguistic, cross-cultural study. *Journal of Pragmatics*, 155, 193–212. <https://doi.org/10.1016/j.pragma.2019.11.001>
- Howe, L. C., MacInnis, B., Krosnick, J. A., Markowitz, E. M., & Socolow, R. (2019). Acknowledging uncertainty impacts public acceptance of climate scientists' predictions. *Nature Climate Change*, 9(11), 863-867.
- Hütter, M., & Ache, F. (2016). Seeking advice: A sampling approach to advice taking. *Judgment and Decision Making*, 11(4), 401-415.
- Izquierdo, E. J., Campbell, C. M., & Goldstone, R. L. (2018, July 7-8). *The Great Melting Pot: generating diversity by combining solutions across a global population*. [Conference Paper]. Collective Intelligence 2018, Zurich, Switzerland.
- Jaquiere, M., & El Zein, M. (2021). Stage 1 Registered Report: How responsibility attributions to self and others relate to outcome ownership in group decisions. *Wellcome Open Research*, 6, 24. <https://doi.org/10.12688/wellcomeopenres.16480.2>
- Ji, L.-J., Zhang, N., Li, Y., Zhang, Z., Harper, G., Khei, M., & Li, J. (2017). Cultural Variations in Reasons for Advice Seeking. *Journal of Behavioral Decision Making*, 30(3), 708–718. <https://doi.org/10.1002/bdm.1995>
- Jin, M., Yang, Z., Freling, T. L., & Janakiraman, N. (2025). The Human Superiority Effect in Advice Taking: A Multimethod Exploration and Implications for Policy Makers and Governmental Organizations. *Journal of Public Policy & Marketing*, 44(3), 350-369. <https://doi.org/10.1177/07439156251320314>
- Johnson, S., Webber, D. J., & Thomas, W. (2007). Which SMEs Use External Business Advice? A Multivariate Subregional Study. *Environment and Planning A: Economy and Space*, 39(8), 1981–1997. <https://doi.org/10.1068/a38327>
- Johnson, T. R., Budescu, D. V., & Wallsten, T. S. (2001). Averaging probability judgments: Monte Carlo analyses of asymptotic diagnostic value. *Journal of Behavioral Decision Making*, 14(2), 123–140. <https://doi.org/10.1002/bdm.369>
- Joiner, T. A., Leveson, L., & Langfield-Smith, K. (2002). Technical Language, Advice Understandability, and Perceptions of Expertise and Trustworthiness: The Case of the Financial Planner. *Australian Journal of Management*, 27(1), 25–43. <https://doi.org/10.1177/031289620202700102>
- Joslyn, S. L., & LeClerc, J. E. (2016). Climate projections and uncertainty communication. *Topics in Cognitive Science*, 8(1), 222-241.
- Jozefowicz, J. (2021). Individual differences in the perception of cue-outcome contingencies: A signal detection analysis. *Behavioural Processes*, 188 (2021), 104398. <https://doi.org/10.1016/j.beproc.2021.104398>

- Julian, J. W., Regula, C. R., & Hollander, E. P. (1968). Effects of prior agreement by others on task confidence and conformity. *Journal of Personality and Social Psychology*, 9(2, Pt.1), 171–178. <https://doi.org/10.1037/h0025851>
- Jungermann, H., & Fischer, K. (2014). Using expertise and experience for giving and taking advice. In *The routines of decision making* (pp. 157-173). Psychology Press.
- Kaanders, P., Sepulveda, P., Folke, T., Ortoleva, P., & De Martino, B. (n.d.). Humans actively sample evidence to support prior beliefs. *eLife*, 11, e71768. <https://doi.org/10.7554/eLife.71768>
- Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, 11(2), 143–157. [https://doi.org/10.1016/0010-0277\(82\)90023-3](https://doi.org/10.1016/0010-0277(82)90023-3)
- Kämmer, J. E., Choshen-Hillel, S., Müller-Trede, J., Black, S. L., & Weibler, J. (2023). A systematic review of empirical studies on advice-based decisions in behavioral and organizational research. *Decision*, 10(2), 107-137.
- Kappes, A., Harvey, A. H., Lohrenz, T., Montague, P. R., & Sharot, T. (2020). Confirmation bias in the utilization of others' opinion strength. *Nature neuroscience*, 23(1), 130-137.
- Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences*, 88(11), 4966-4970. <https://doi.org/10.1073/pnas.88.11.4966>
- Kassirer, S., Levine, E. E., & Gaertig, C. (2020). Decisional autonomy undermines advisees' judgments of experts in medicine and in life. *Proceedings of the National Academy of Sciences*, 117(21), 11368-11378. <https://doi.org/10.1073/pnas.1910572117>
- Kassirer, S., Levine, E., & Gaertig, C. (2019). *Decisional Autonomy Undermines Advisees' Judgments of Experts In Medicine And In Life* (SSRN Scholarly Paper No. 3455418). Social Science Research Network. <https://doi.org/10.2139/ssrn.3455418>
- Kausel, E. E., Culbertson, S. S., Leiva, P. I., Slaughter, J. E., & Jackson, A. T. (2015). Too arrogant for their own good? Why and when narcissists dismiss advice. *Organizational Behavior and Human Decision Processes*, 131, 33–50. <https://doi.org/10.1016/j.obhdp.2015.07.006>
- Kennedy, M. M. (1987). Inexact Sciences: Professional Education and the Development of Expertise. Issue Paper 87-2.
- Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1322-1337. <https://doi.org/10.1098/rstb.2012.0037>
- King, A. J., Cheng, L., Starke, S. D., & Myatt, J. P. (2012). Is the true 'wisdom of the crowd' to copy successful individuals?. *Biology letters*, 8(2), 197-200. <https://doi.org/10.1098/rsbl.2011.0795>
- Klayman, J. (1984). Learning from feedback in probabilistic environments. *Acta Psychologica*, 56(1-3), 81-92. [https://doi.org/10.1016/0001-6918\(84\)90009-X](https://doi.org/10.1016/0001-6918(84)90009-X)
- Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational behavior and human decision processes*, 79(3), 216-247. <https://doi.org/10.1006/obhd.1999.2847>
- Kooiman, B. E., Robberegt, S. J., Albers, C. J., Bockting, C. L., Stikkelbroek, Y. A., & Nauta, M. H. (2023). Congruency of multimodal data-driven personalization with shared decision-making for StayFine: Individualized app-based relapse

- prevention for anxiety and depression in young people. *Frontiers in Psychiatry*, 14, 1229713.
- Kramer, M. M. (2016). Financial literacy, confidence and financial advice seeking. *Journal of Economic Behavior & Organization*, 131, 198–217. <https://doi.org/10.1016/j.jebo.2016.08.016>
- Kronrod, A., Grinstein, A., & Wathieu, L. (2012). Go Green! Should Environmental Messages be So Assertive? *Journal of Marketing*, 76(1), 95-102. <https://doi.org/10.1509/jm.10.0416>
- Kronrod, A., Grinstein, A., & Wathieu, L. (2025). When Needs Don't Translate into Wants: Choosing the Right Tone to Communicate About Healthy Behaviors. *Health Communication*, 40(6), 980–991. <https://doi.org/10.1080/10410236.2024.2383800>
- Kronrod, A., Grinstein, A., & Wathieu, L. (2025). When Needs Don't Translate into Wants: Choosing the Right Tone to Communicate About Healthy Behaviors. *Health Communication*, 40(6), 980–991. <https://doi.org/10.1080/10410236.2024.2383800>
- Krueger, J. I. (2003). Return of the ego--Self-referent information as a filter for social prediction: Comment on Karniol (2003).
- Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(5), 1083-1119.
- Kvam, P. D., & Hintze, A. (2018). Rewards, risks, and reaching the right strategy: Evolutionary paths from heuristics to optimal decisions. *Evolutionary Behavioral Sciences*, 12(3), 177–190. <https://doi.org/10.1037/ebs0000115>
- Kyburg Jr, H. E. (1987). Bayesian and non-Bayesian evidential updating. *Artificial intelligence*, 31(3), 271-293. [https://doi.org/10.1016/0004-3702\(87\)90068-3](https://doi.org/10.1016/0004-3702(87)90068-3)
- Laage-Thomsen, J. (2022). Professional expertise in policy advisory systems: How administrators and consultants built behavioral insights in Danish public agencies. *Public Administration*, 100(3), 472–487. <https://doi.org/10.1111/padm.12725>
- Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple-cue learning. *Journal of Experimental Psychology: General*, 135(2), 162–183. <https://doi.org/10.1037/0096-3445.135.2.162>
- Lazer, D., & Friedman, A. (2007). The network structure of exploration and exploitation. *Administrative science quarterly*, 52(4), 667-694. <https://doi.org/10.2189/asqu.52.4.667>
- Lee, J., Rajtmajer, S., Srivatsavaya, E., & Wilson, S. (2023). Online self-disclosure, social support, and user engagement during the COVID-19 pandemic. *ACM Transactions on Social Computing*, 6(3-4), 1-31. <https://doi.org/10.1145/3617654>
- Lee, M. D., & Dry, M. J. (2006). Decision making and confidence given uncertain advice. *Cognitive Science*, 30(6), 1081-1095. [https://doi.org/10.1207/s15516709cog0000\\_71](https://doi.org/10.1207/s15516709cog0000_71)
- Lei, X., Wu, H., Deng, Z., & Ye, Q. (2023). Self-disclosure, social support and postpartum depressive mood in online social networks: a social penetration theory perspective. *Information Technology & People*, 36(1), 433-453.
- Li, J., Li, M., Sun, Y., Zhang, G., Fan, W., & Zhong, Y. (2024). The impact of social hierarchies on neural response to feedback evaluations after advice giving. *Human Brain Mapping*, 45(2), e26611.
- Liang, F., Qian, C., & Duan, J. (2023). Why and when does shyness hinder people from seeking advice? *Journal of Pacific Rim Psychology*, 17, 18344909231154927. <https://doi.org/10.1177/18344909231154927>

- Liang, H., & Xue, Y. (2022). Save Face or Save Life: Physicians' Dilemma in Using Clinical Decision Support Systems. *Information Systems Research*, 33(2), 737–758. <https://doi.org/10.1287/isre.2021.1082>
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge University Press. <https://doi.org/10.1017/CBO9780511809477.023>
- Lim, J. S., & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: Its effectiveness and biases. *Journal of Behavioral Decision Making*, 8(3), 149–168. <https://doi.org/10.1002/bdm.3960080302>
- Linnainmaa, J. T., Melzer, B. T., & Previtro, A. (2021). The misguided beliefs of financial advisors. *The Journal of Finance*, 76(2), 587–621.
- Little, D. R., & Lewandowsky, S. (2009). Better learning with more error: Probabilistic feedback increases sensitivity to correlated cues in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(4), 1041–1061. <https://doi.org/10.1037/a0015902>
- Locander, W. B., & Hermann, P. W. (1979). The effect of self-confidence and anxiety on information seeking in consumer risk reduction. *Journal of Marketing Research*, 16(2), 268–274. <https://doi.org/10.1177/002224377901600211>
- Løhre, E., & Halkjelsvik, T. (2024). Advice taking when the stakes are high: Evidence from a game show. *Judgment and Decision Making*, 19, e25. <https://doi.org/10.1017/jdm.2024.4>
- Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the national academy of sciences*, 108(22), 9020–9025.
- Loschelder, D. D., Friese, M., Schaerer, M., & Galinsky, A. D. (2016). The too-much-precision effect: When and why precise anchors backfire with experts. *Psychological science*, 27(12), 1573–1587. <https://doi.org/10.1177/09567976166666074>
- Ma, S., Kor, Y. Y., & Seidl, D. (2020). CEO Advice Seeking: An Integrative Framework and Future Research Agenda. *Journal of Management*, 46(6), 771–805. <https://doi.org/10.1177/0149206319885430>
- MacGeorge, E. L., & Swol, L. M. V. (2018). *The Oxford Handbook of Advice*. Oxford University Press.
- MacGeorge, E. L., Guntzviller, L. M., Hanasono, L. K., & Feng, B. (2008). “Good” and “bad” advice: How to advise more effectively. In M. T. Motley (Ed.), *Studies in applied interpersonal communication* (pp. 145–164). Sage Publications.
- MacGeorge, E. L., Lichtman, R. M., & Pressey, L. C. (2002). The evaluation of advice in supportive interactions: Facework and contextual factors. *Human Communication Research*, 28(3), 451–463.
- Mahmoodi, A., Bahrami, B., & Mehring, C. (2018). Reciprocity of social influence. *Nature communications*, 9(1), 2474. [10.1038/s41467-018-04925-y](https://doi.org/10.1038/s41467-018-04925-y)
- Manian, S., & Sheth, K. (2021). Follow my lead: Assertive cheap talk and the gender gap. *Management Science*, 67(11), 6880–6896.
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization science*, 2(1), 71–87. <https://doi.org/10.1287/orsc.2.1.71>
- Mason, W., & Watts, D. J. (2012). Collaborative learning in networks. *Proceedings of the National Academy of Sciences*, 109(3), 764–769. <https://doi.org/10.1073/pnas.1110069108>

- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model Of Organizational Trust. *Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.5465/amr.1995.9508080335>
- McGarty, C., Turner, J. C., Oakes, P. J., & Haslam, S. A. (1993). The creation of uncertainty in the influence process: The roles of stimulus information and disagreement with similar others. *European Journal of Social Psychology*, 23(1), 17–38. <https://doi.org/10.1002/ejsp.2420230103>
- McKee, S. P., & Westhe, G. (1978). Improvement in vernier acuity with practice. *Perception & psychophysics*, 24(3), 258-262. <https://doi.org/10.3758/BF03206097>
- Mercier, H., Yama, H., Kawasaki, Y., Adachi, K., & Van der Henst, J.-B. (2012). Is the use of averaging in advice taking modulated by culture? *Journal of Cognition and Culture*, 12(1–2), 1–16. <https://doi.org/10.1163/156853712X633893>
- Meshi, D., Biele, G., Korn, C. W., & Heekeren, H. R. (2012). How expert advice influences decision making. *PLoS One*, 7(11), e49748.
- Meyer, C., Ulbricht, S., Baumeister, S. E., Schumann, A., Rüge, J., Bischof, G., ... & John, U. (2008). Proactive interventions for smoking cessation in general medical practice: a quasi-randomized controlled trial to examine the efficacy of computer-tailored letters and physician-delivered brief advice. *Addiction*, 103(2), 294-304.
- Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian probability: From neural origins to behavior. *Neuron*, 88(1), 78-92. <http://dx.doi.org/10.1016/j.neuron.2015.09.039>
- Milyavsky, M., & Gvili, Y. (2024). Advice-taking vs. combining opinions: Framing social information as advice increases source's perceived helping intentions, trust, and influence. *Organizational Behavior and Human Decision Processes*, 183, 104328. <https://doi.org/10.1016/j.obhdp.2024.104328>
- Milyavsky, M., Kruglanski, A. W., Chernikova, M., & Schori-Eyal, N. (2017). Evidence for arrogance: On the relative importance of expertise, outcome, and manner. *PloS one*, 12(7), e0180420.
- Minson, J. A., Liberman, V., & Ross, L. (2011). Two to tango: Effects of collaboration and disagreement on dyadic judgment. *Personality & Social Psychology Bulletin*, 37(10), 1325–1338. <https://doi.org/10.1177/0146167211410436>
- Mole, K., North, D., & Baldock, R. (2017). Which SMEs seek external support? Business characteristics, management behaviour and external influences in a contingency approach. *Environment and Planning C: Politics and Space*, 35(3), 476–499. <https://doi.org/10.1177/0263774X16665362>
- Molleman, L., Tump, A. N., Gradassi, A., Herzog, S., Jayles, B., Kurvers, R. H. J. M., & van den Bos, W. (2020). Strategies for integrating disparate social information. *Proceedings of the Royal Society B: Biological Sciences*, 287(1939), 20202413. <https://doi.org/10.1098/rspb.2020.2413>
- Montmarquette, C., & Viennot-Briot, N. (2015). The value of financial advice. *Annals of Economics and Finance*, 16(1), 69-94.
- Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes factors for common designs* (R package version 0.9.12-4.2). <https://CRAN.R-project.org/package=BayesFactor>
- Morgan, T. J., Laland, K. N., & Harris, P. L. (2015). The development of adaptive conformity in young children: effects of uncertainty and consensus. *Developmental science*, 18(4), 511-524.

- Moscovici, S., & Zavalloni, M. (1969). The group as a polarizer of attitudes. *Journal of Personality and Social Psychology*, 12(2), 125–135. <https://doi.org/10.1037/h0027568>
- Motley, M. T. (2008). *Studies in Applied Interpersonal Communication*. SAGE Publications.
- Murphy, G. L., & Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: what a superordinate is good for. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(4), 572-586.
- Mynatt, C., & Sherman, S. J. (1975). Responsibility attribution in groups and individuals: A direct test of the diffusion of responsibility hypothesis. *Journal of Personality and Social Psychology*, 32(6), 1111–1118. <https://doi.org/10.1037/0022-3514.32.6.1111>
- Nassar, M. R., Wilson, R. C., Heasley, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience*, 30(37), 12366-12378. <https://doi.org/10.1523/JNEUROSCI.0822-10.2010>
- Navajas, J., Hindocha, C., Foda, H., Keramati, M., Latham, P. E., & Bahrami, B. (2017). The idiosyncratic nature of confidence. *Nature Human Behaviour*, 1(11), 810–818. <https://doi.org/10.1038/s41562-017-0215-1>
- Newell, B. R., Weston, N. J., Tunney, R. J., & Shanks, D. R. (2009). The effectiveness of feedback in multiple-cue probability learning. *Quarterly Journal of Experimental Psychology*, 62(5), 890-908. <https://doi.org/10.1080/17470210802351411>
- Nückles, Matthias, Jörg Wittwer, and Alexander Renkl. "Information about a layperson's knowledge supports experts in giving effective and efficient online advice to laypersons." *Journal of Experimental Psychology: Applied* 11, no. 4 (2005): 219.
- Nyman, S. R., & Yardley, L. (2009). Web-site-based tailored advice to promote strength and balance training: an experimental evaluation. *Journal of aging and physical activity*, 17(2), 210-222.
- Ollesch, H., Heineken, E., & Schulte, F. P. (2006). *Physical or Virtual Presence of the Experimenter: Psychological Online-Experiments in Different Settings*.
- Önkal, D., Sinan Gönül, M., Goodwin, P., Thomson, M., & Öz, E. (2017). Evaluating expert advice in forecasting: Users' reactions to presumed vs. experienced credibility. *International Journal of Forecasting*, 33(1), 280–297. <https://doi.org/10.1016/j.ijforecast.2015.12.009>
- Osborne, M. R., & Bailey, E. R. (2025). Me vs. the machine? Subjective evaluations of human- and AI-generated advice. *Scientific Reports*, 15(1), 3980. <https://doi.org/10.1038/s41598-025-86623-6>
- Paik, J. E. (2020). The contextual effects of advice solicitation on advice outcomes: The role of perceived face threats and psychological reactance. *Communication Monographs*, 87(1), 70-91.
- Palmeira, M. (2020). Advice in the presence of external cues: The impact of conflicting judgments on perceptions of expertise. *Organizational Behavior and Human Decision Processes*, 156, 82-96. <https://doi.org/10.1016/j.obhdp.2019.11.009>
- Palmeira, M., & Romero Lopez, M. (2023). The opposing impacts of advice use on perceptions of competence. *Journal of Behavioral Decision Making*, 36(4), e2318. <https://doi.org/10.1002/bdm.2318>
- Palmeira, M., & Spassova, G. (2024). Consulting Multiple Advisors: When It Hurts and When It Does Not Hurt the Advisor–Advisee Relationship?. *Journal of Behavioral Decision Making*, 37(2), e2382.

- Pan, W., Feng, B., & Skye Wingate, V. (2018). What you say is what you get: How self-disclosure in support seeking affects language use in support provision in online support forums. *Journal of Language and Social Psychology, 37*(1), 3-27.
- Patt, A. G., Bowles, H. R., & Cash, D. W. (2006). Mechanisms for enhancing the credibility of an adviser: Prepayment and aligned incentives. *Journal of Behavioral Decision Making, 19*(4), 347-359.
- Pawitan, Y., & Lee, Y. (2021). Confidence as likelihood. *Statistical Science, 36*(4), 509-517.
- Peng, Z. (2022). An epistemic illumination of the acceptance of advice in Chinese phone-in counseling for family problems. *Discourse Studies, 24*(3), 291-306.
- Pereg, M., Hertz, U., Ben-Artzi, I., & Shahar, N. (2024). Disentangling the contribution of individual and social learning processes in human advice-taking behavior. *Npj Science of Learning, 9*(1), 4. <https://doi.org/10.1038/s41539-024-00214-0>
- Perry, P. (2011). Concept Analysis: Confidence/Self-confidence. *Nursing Forum, 46*(4), 218–230. <https://doi.org/10.1111/j.1744-6198.2011.00230.x>
- Pescetelli, N., & Yeung, N. (2021). The role of decision confidence in advice-taking and trust formation. *Journal of Experimental Psychology: General, 150*(3), 507–526. <https://doi.org/10.1037/xge0000960>
- Pescetelli, N., & Yeung, N. (2022). *The effects of social communication dynamics on opinion change* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/kmuqw>
- Pescetelli, N., Hauperich, A. K., & Yeung, N. (2021). Confidence, advice seeking and changes of mind in decision making. *Cognition, 215*, 104810. <https://doi.org/10.1016/j.cognition.2021.104810>
- Phillips, J. M. (1999). Antecedents of Leader Utilization of Staff Input in Decision-Making Teams. *Organizational Behavior and Human Decision Processes, 77*(3), 215–242. <https://doi.org/10.1006/obhd.1998.2819>
- Polyanskaya, L. (2023). I know that I know. But do I know that I do not know? *Frontiers in Psychology, 14*. <https://doi.org/10.3389/fpsyg.2023.1128200>
- Poston, R. S., Akbulut, A. Y., & Looney, C. A. (2005). Online advice taking: Examining the effects of self-efficacy, computerized sources, and perceived credibility. *SIGHCI 2005 Proceedings, 5*.
- Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: Distinct probabilistic quantities for different goals. *Nature Neuroscience, 19*(3), 366–374. <https://doi.org/10.1038/nn.4240>
- Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making, 17*(1), 39-57. <https://doi.org/10.1002/bdm.460>
- Prissé, B., & Jorrat, D. (2022). Lab vs online experiments: No differences. *Journal of Behavioral and Experimental Economics, 100*, 101910. <https://doi.org/10.1016/j.socec.2022.101910>
- Pulford, B. D., Colman, A. M., Buabang, E. K., & Krockow, E. M. (2018). The persuasive power of knowledge: Testing the confidence heuristic. *Journal of Experimental Psychology: General, 147*(10), 1431-1444. <https://doi.org/10.1037/xge0000471>
- Rader, C. A., Larrick, R. P., & Soll, J. B. (2017). Advice as a form of social influence: Informational motives and the consequences for accuracy. *Social and Personality Psychology Compass, 11*(8), e12329. <https://doi.org/10.1111/spc3.12329>
- Rebholz, T. R., & Hütter, M. (2022). The advice less taken: The consequences of receiving unexpected advice. *Judgment and Decision Making, 17*(4), 816-848.

- Reips, U.-D. (2001). The Web Experimental Psychology Lab: Five years of data collection on the Internet. *Behavior Research Methods, Instruments, & Computers*, 33(2), 201–211. <https://doi.org/10.3758/BF03195366>
- Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., ... & Laland, K. N. (2010). Why copy others? Insights from the social learning strategies tournament. *Science*, 328(5975), 208–213. <https://doi.org/10.1126/science.1184719>
- Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J., & Fleming, S. M. (2020). Confidence drives a neural confirmation bias. *Nature communications*, 11(1), 2634. <https://doi.org/10.1038/s41467-020-16278-6>
- Rose, S. A., Poynter, P. S., Anderson, J. W., Noar, S. M., & Conigliaro, J. (2013). Physician weight loss advice and patient weight loss behavior change: a literature review and meta-analysis of survey data. *International journal of obesity*, 37(1), 118–128. <https://doi.org/10.1038/ijo.2012.24>
- Rosenthal, D. E. (1974). *Lawyer and Client: Who's in Charge*. Russell Sage Foundation. <https://doi.org/10.2307/2148900>
- Rotenstreich, N. (1972). On confidence. *Philosophy*, 47(182), 348–358
- Sah, S., Moore, D. A., & MacCoun, R. J. (2013). Cheap talk and credibility: The consequences of confidence and accuracy on advisor credibility and persuasiveness. *Organizational Behavior and Human Decision Processes*, 121(2), 246–255. <https://doi.org/10.1016/j.obhdp.2013.02.001>
- Salacuse, J. (2018). Advice in Government and Policy Making. In E. L. MacGeorge & L. M. Van Swol (Eds), *The Oxford Handbook of Advice* (pp. 321–342). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190630188.013.16>
- Sauter, M., Stefani, M., & Mack, W. (2022). Equal Quality for Online and Lab Data: A Direct Comparison from Two Dual-Task Paradigms. *Open Psychology*, 4(1), 47–59. <https://doi.org/10.1515/psych-2022-0003>
- Schidelko, L. P., Schünemann, B., Rakoczy, H., & Proft, M. (2021). Online Testing Yields the Same Results as Lab Testing: A Validation Study With the False Belief Task. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.703238>
- Schrah, G. E., Dalal, R. S., & Sniezek, J. A. (2006). No decision-maker is an island: Integrating expert advice with information acquisition. *Journal of Behavioral Decision Making*, 19(1), 43–60. <https://doi.org/10.1002/bdm.514>
- Schultze, T., & Chen, Z. (2025). The development of a task to study advice taking across nations and its application in a China-Germany comparison. *Judgment and Decision Making*, 20, e8. <https://doi.org/10.1017/jdm.2024.43>
- Schultze, T., & Loschelder, D. D. (2020). How numeric advice precision affects advice taking. *Journal of Behavioral Decision Making*, 34(3), 303–310. <https://doi.org/10.1002/bdm.2211>
- Schultze, T., Rakotoarisoa, A. F., & Stefan, S. H. (2015). Effects of distance between initial estimates and advice on advice utilization. *Judgment and Decision making*, 10(2), 144–171. <https://doi.org/10.1017/S1930297500003922>
- Sciandra, M. R. (n.d.). Beyond closed doors: Examining consumer reactions to public and private advice. *Journal of Marketing Communications*, 0(0), 1–22. <https://doi.org/10.1080/13527266.2024.2419027>
- See, K. E., Morrison, E. W., Rothman, N. B., & Soll, J. B. (2011). The detrimental effects of power on confidence, advice taking, and accuracy. *Organizational behavior and human decision processes*, 116(2), 272–285. <https://doi.org/10.1016/j.obhdp.2011.07.006>

- Seppänen, J. T., Forsman, J. T., Mönkkönen, M., & Thomson, R. L. (2007). Social information use is a process across time, space, and ecology, reaching heterospecifics. *Ecology*, 88(7), 1622-1633. <https://doi.org/10.1890/06-1757.1>
- Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, 134(2), 207–222. <https://doi.org/10.1037/0033-2909.134.2.207>
- Shanteau, J., & Stewart, T. R. (1992). Why study expert decision making? Some historical perspectives and comments. *Organizational Behavior and Human Decision Processes*, 53(2), 95-106. [https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1016/0749-5978\(92\)90057-E](https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1016/0749-5978(92)90057-E)
- Sherf, E. N., & Morrison, E. W. (2020). I do not need feedback! Or do I? Self-efficacy, perspective taking, and feedback seeking. *Journal of Applied Psychology*, 105(2), 146–165. <https://doi.org/10.1037/apl0000432>
- Siedlecka, M., Skóra, Z., Paulewicz, B., Fijałkowska, S., Timmermans, B., & Wierchoń, M. (2019). Responses improve the accuracy of confidence judgements in memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(4), 712–723. <https://doi.org/10.1037/xlm0000608>
- Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-adviser decision making. *Organizational behavior and human decision processes*, 62(2), 159-174. <https://doi.org/10.1006/obhd.1995.1040>
- Sniezek, J. A., & Van Swol, L. M. (2001). Trust, Confidence, and Expertise in a Judge-Advisor System. *Organizational Behavior and Human Decision Processes*, 84(2), 288–307. <https://doi.org/10.1006/obhd.2000.2926>
- Sniezek, J. A., Schrah, G. E., & Dalal, R. S. (2004). Improving judgement with prepaid expert advice. *Journal of Behavioral Decision Making*, 17(3), 173–190. <https://doi.org/10.1002/bdm.468>
- Snijders, C., Conijn, R., de Fouw, E., & van Berlo, K. (2022). Humans and algorithms detecting fake news: Effects of individual and contextual confidence on trust in algorithmic advice. *International Journal of Human-Computer Interaction*. Advance online publication. <https://doi.org/10.1080/10447318.2022.2097601>
- Snijders, C., Conijn, R., De Fouw, E., & Van Berlo, K. (2023). Humans and Algorithms Detecting Fake News: Effects of Individual and Contextual Confidence on Trust in Algorithmic Advice. *International Journal of Human-Computer Interaction*, 39(7), 1483–1494. <https://doi.org/10.1080/10447318.2022.2097601>
- Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 780–805. <https://doi.org/10.1037/a0015145>
- Soll, J. B., & Mannes, A. E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting*, 27(1), 81-102. <https://doi.org/10.1016/j.ijforecast.2010.05.003>
- Soll, J. B., Palley, A. B., & Rader, C. A. (2021). The bad thing about good advice: Understanding when and how advice exacerbates overconfidence. *Management Science*, 68(4), 2949-2969. <https://doi.org/10.1287/mnsc.2021.3987>
- Soll, J. B., Palley, A. B., & Rader, C. A. (2022). The Bad Thing About Good Advice: Understanding When and How Advice Exacerbates Overconfidence. *Management Science*, 68(4), 2949–2969. <https://doi.org/10.1287/mnsc.2021.3987>

- Soltani, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews Neuroscience*, 20(10), 635-644. <https://doi.org/10.1038/s41583-019-0180-y>
- Stanciu, O., & Fiser, J. (2022). Do humans recalibrate the confidence of advisers or take their confidence at face value? *Proceedings of the Annual Meeting of the Cognitive Science Society*, 44(44), 3432-3438. <https://escholarship.org/uc/item/94b1q5hq>
- Stavrova, O., & Evans, A. M. (2019). Examining the trade-off between confidence and optimism in future forecasts. *Journal of Behavioral Decision Making*, 32(1), 3–14. <https://doi.org/10.1002/bdm.2087>
- Stewart, T. R., Mumpower, J. L., & James Holzworth, R. (2012). Learning to make selection and detection decisions: The roles of base rate and feedback. *Journal of Behavioral Decision Making*, 25(5), 522-533. <https://doi.org/10.1002/bdm.755>
- Strong, S. R. (1968). Counseling: An interpersonal influence process. *Journal of counseling Psychology*, 15(3), 215-224. <https://doi.org/10.1037/h0020229>
- Sun, T., Horn, M., & Merritt, D. (2004). Values and lifestyles of individualists and collectivists: A study on Chinese, Japanese, British and US consumers. *Journal of Consumer Marketing*, 21(5), 318–331. <https://doi.org/10.1108/07363760410549140>
- Sunder, A., Palmer, L., Chatterjee, S., & Goetz, J. (2024). Benefits of consistent and comprehensive financial advice during the Great Recession. *Journal of Behavioral and Experimental Finance*, 41, 100900. <https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1016/j.jbef.2024.100900>
- Surowiecki J (2004) *The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations* (Doubleday Books, New York).
- Sutton, R. S., & Barto, A. G. (1999). Reinforcement learning. *Journal of Cognitive Neuroscience*, 11(1), 126-134.
- Tan, G. K. S. (2020). Robo-advisors and the financialization of lay investors. *Geoforum*, 117, 46-60. <https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1016/j.geoforum.2020.09.004>
- Tannenbaum, D., Fox, C. R., & Ülkümen, G. (2017). Judgment extremity and accuracy under epistemic vs. aleatory uncertainty. *Management Science*, 63(2), 497-518. <https://doi.org/10.1287/mnsc.2015.2344>
- Tenney, E. R., & Spellman, B. A. (2011). Complex Social Consequences of Self-Knowledge. *Social Psychological and Personality Science*, 2(4), 343–350. <https://doi.org/10.1177/1948550610390965>
- Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science*, 18(1), 46-50. <https://doi.org/10.1111/j.1467-9280.2007.01847.x>
- Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge University Press. <https://doi.org/10.1017/CBO9780511809477.023>
- Tenney, E. R., Meikle, N. L., Hunsaker, D., Moore, D. A., & Anderson, C. (2019). Is overconfidence a social liability? The effect of verbal versus nonverbal expressions of confidence. *Journal of Personality and Social Psychology*, 116(3), 396–415. <https://doi.org/10.1037/pspi0000150>

- Tenney, E. R., Spellman, B. A., & MacCoun, R. J. (2008). The benefits of knowing what you know (and what you don't): How calibration affects credibility. *Journal of Experimental Social Psychology*, 44(5), 1368-1375. <https://doi.org/10.1016/j.jesp.2008.04.006>
- Thomas, J. P., & McFadyen, R. G. (1995). The confidence heuristic: A game-theoretic analysis. *Journal of Economic Psychology*, 16(1), 97-113. [https://doi.org/10.1016/0167-4870\(94\)00032-6](https://doi.org/10.1016/0167-4870(94)00032-6)
- Toelch, U., Bruce, M. J., Newson, L., Richerson, P. J., & Reader, S. M. (2014). Individual consistency and flexibility in human social information use. *Proceedings of the Royal Society B: Biological Sciences*, 281(1776), 20132864. <https://doi.org/10.1098/rspb.2013.2864>
- Tost, L. P., Gino, F., & Larrick, R. P. (2012). Power, competitiveness, and advice taking: Why the powerful don't listen. *Organizational behavior and human decision processes*, 117(1), 53-65. <https://doi.org/10.1016/j.obhdp.2011.10.001>
- Trafimow, D., & Sniezek, J. A. (1994). Perceived expertise and its effect on confidence. *Organizational behavior and human decision processes*, 57(2), 290-302. <https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1006/obhd.1994.1016>
- Treffenstädt, C. (2017). *The Consequences of Disregarding Advice in Judge-Advisor Systems* (Doctoral dissertation, Niedersächsische Staats- und Universitätsbibliothek Göttingen).
- Uittenhove, K., Jeanneret, S., & Vergauwe, E. (n.d.). From Lab-Testing to Web-Testing in Cognitive Research: Who You Test is More Important than how You Test. *Journal of Cognition*, 6(1), 13. <https://doi.org/10.5334/joc.259>
- Ülkümen, G., Fox, C. R., & Malle, B. F. (2016). Two dimensions of subjective uncertainty: Clues from natural language. *Journal of Experimental Psychology: General*, 145(10), 1280–1297. <http://dx.doi.org/10.1037/xge0000202>
- Valone, T. J. (2007). From eavesdropping on performance to copying the behavior of others: a review of public information use. *Behavioral ecology and sociobiology*, 62, 1-14. <http://dx.doi.org/10.1007/s00265-007-0439-6>
- Van Der Bles, A. M., van der Linden, S., Freeman, A. L., & Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences*, 117(14), 7672-7683. <https://doi.org/10.1073/pnas.1913678117>
- Van Swol, L. M. (2009). Discussion and perception of information in groups and judge-advisor systems. *Communication Monographs*, 76(1), 99-120. <https://doi.org/10.1080/03637750802378781>
- Van Swol, L. M. (2011). Forecasting another's enjoyment versus giving the right answer: Trust, shared values, task effects, and confidence in improving the acceptance of advice. *International Journal of Forecasting*, 27(1), 103-120. <https://doi.org/10.1016/j.ijforecast.2010.03.002>
- Van Swol, L. M., & Sniezek, J. A. (2005). Factors affecting the acceptance of expert advice. *British journal of social psychology*, 44(3), 443-461. <https://doi.org/10.1348/014466604X17092>
- Van Swol, L. M., MacGeorge, E. L., & Prah, A. (2017). Advise with permission?: The effects of advice solicitation on advice outcomes. *Communication Studies*, 68(4), 476-492. <https://doi.org/10.1080/10510974.2017.1363795>

- Van Zant, A. B. (2022). Strategically overconfident (to a fault): How self-promotion motivates advisor confidence. *Journal of Applied Psychology*, *107*(1), 109-129. <https://doi.org/10.1037/apl0000879>
- Vasalou, A., Hopfensitz, A., & Pitt, J. V. (2008). In praise of forgiveness: Ways for repairing trust breakdowns in one-off online interactions. *International Journal of Human-Computer Studies*, *66*(6), 466-480. <https://doi.org/10.1016/j.ijhcs.2008.02.001>
- Veenman, M. V., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and learning*, *1*, 3-14. <https://doi.org/10.1007/s11409-006-6893-0>
- Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, *36*, 1-48. <https://doi.org/10.18637/jss.v036.i03>
- Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, *36*, 1-48. <https://doi.org/10.18637/jss.v036.i03>
- Vodrahalli, K., Daneshjou, R., Gerstenberg, T., & Zou, J. (2022). Do Humans Trust Advice More if it Comes from AI? An Analysis of Human-AI Interactions. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 763-777. <https://doi.org/10.1145/3514094.3534150>
- Vullioud, C., Clément, F., Scott-Phillips, T., & Mercier, H. (2017). Confidence as an expression of commitment: Why misplaced expressions of confidence backfire. *Evolution and Human Behavior*, *38*(1), 9-17. <https://doi.org/10.1016/j.evolhumbehav.2016.06.002>
- Walters, D. J., Ülkümen, G., Tannenbaum, D., Erner, C., & Fox, C. R. (2023). Investor behavior under epistemic vs. aleatory uncertainty. *Management Science*, *69*(5), 2761-2777. <https://doi.org/10.1287/mnsc.2022.4489>
- Walther, J. B., Loh, T., & Granka, L. (2005). Let me count the ways: The interchange of verbal and nonverbal cues in computer-mediated and face-to-face affinity. *Journal of language and social psychology*, *24*(1), 36-65. <https://doi.org/10.1177/0261927X04273036>
- Wang, X., & Du, X. (2018). Why does advice discounting occur? The combined roles of confidence and trust. *Frontiers in psychology*, *9*, 2381. <https://doi.org/10.3389/fpsyg.2018.02381>
- Wang, X., Huang, X., & Zhang, E. (2025). How individuals evaluate the confidence of advice from advisors with high-and low-status: A behavioural and ERP study. *Biological Psychology*, *194*, 108978. <https://doi.org/10.1016/j.biopsycho.2024.108978>
- Wang, Z., Wei, L., & Xue, L. (2024). Overcoming medical overuse with ai assistance: An experimental investigation. *arXiv preprint arXiv:2405.10539*. <https://doi.org/10.1016/j.jhealeco.2025.103043>
- Watson, R., & Morgan, T. J. (2025). An experimental test of epistemic vigilance: Competitive incentives increase dishonesty and reduce social influence. *Cognition*, *257*, 106066. <https://doi.org/10.1016/j.cognition.2025.106066>
- Wei, W., Cheong, C. M., Zhu, X., & Lu, Q. (2024). Comparing self-reflection and peer feedback practices in an academic writing task: a student self-efficacy perspective. *Teaching in Higher Education*, *29*(4), 896-912. <https://doi.org/10.1080/13562517.2022.2042242>

- Yaniv, I. (1997). Weighting and trimming: Heuristics for aggregating judgments under uncertainty. *Organizational behavior and human decision processes*, 69(3), 237-249. <https://doi.org/10.1006/obhd.1997.2685>
- Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, 93(1), 1-13. <https://doi.org/10.1016/j.obhdp.2003.08.002>
- Yaniv, I., & Choshen-Hillel, S. (2012). Exploiting the Wisdom of Others to Make Better Decisions: Suspending Judgment Reduces Egocentrism and Increases Accuracy. *Journal of Behavioral Decision Making*, 25(5), 427-434. <https://doi.org/10.1002/bdm.740>
- Yaniv, I., & Kleinberger, E. (2000). Advice-taking in decision making: Egocentric discounting and reputation formation. *Organizational behavior and human decision processes*, 83(2), 260-281. <https://doi.org/10.1006/obhd.2000.2909>
- Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes*, 103(1), 104-120. <https://doi.org/10.1016/j.obhdp.2006.05.006>
- Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2011). Receiving advice on matters of taste: Similarity, majority influence, and taste discrimination. *Organizational Behavior and Human Decision Processes*, 115(1), 111-120. <https://doi.org.ezproxy-prd.bodleian.ox.ac.uk/10.1016/j.obhdp.2010.11.006>
- Yardley, L., & Nyman, S. R. (2007). Internet provision of tailored advice on falls prevention activities for older people: a randomized controlled evaluation. *Health promotion international*, 22(2), 122-128. <https://doi.org/10.1093/heapro/dam007>
- Yates, J. F., & De Oliveira, S. (2016). Culture and decision making. *Organizational behavior and human decision processes*, 136, 106-118. <https://doi.org/10.1016/j.obhdp.2016.05.003>
- Yates, J. F., Price, P. C., Lee, J. W., & Ramirez, J. (1996). Good probabilistic forecasters: The 'consumer's' perspective. *International Journal of Forecasting*, 12(1), 41-56. [https://doi.org/10.1016/0169-2070\(95\)00636-2](https://doi.org/10.1016/0169-2070(95)00636-2)
- Yitshaki, R. (2024). Advice seeking and mentors' influence on entrepreneurs' role identity and business-model change. *Journal of Small Business Management*, 63(2), 1-41. <https://doi.org/10.1080/00472778.2024.2307494>
- Yitshaki, R. (2025). Advice seeking and mentors' influence on entrepreneurs' role identity and business-model change. *Journal of Small Business Management*, 63(1), 70-110. <https://doi.org/10.1080/00472778.2024.2307494>
- Yoon, H., Scopelliti, I., & Morewedge, C. K. (2021). Decision making can be improved through observational learning. *Organizational Behavior and Human Decision Processes*, 162, 155-188. <https://doi.org/10.1016/j.obhdp.2020.10.011>
- Zaffora Blando, F. (2022). Bayesian merging of opinions and algorithmic randomness. *The British Journal for the Philosophy of Science*. <https://doi.org/10.1086/721758>
- Zell, E., Strickhouser, J. E., Sedikides, C., & Alicke, M. D. (2020). The better-than-average effect in comparative self-evaluation: A comprehensive review and meta-analysis. *Psychological Bulletin*, 146(2), 118-149. <https://doi.org/10.1037/bul0000218>
- Zhang, J., Balaji, M. S., Luo, J., Pandey, N., & Wirtz, J. (2025). In diversity lies opportunity: complementary product recommendations, message framing and cross-selling. *Internet Research*. Vol. ahead-of-print No. ahead-of-print. <https://doi.org/10.1108/INTR-05-2024-0789>

- Zhang, J., Cai, L., & Zheng, X. (2024). How influencers can enhance consumer responses by value co-creation. *European Journal of Marketing*, 58(11), 2566–2595. <https://doi.org/10.1108/EJM-05-2023-0362>
- Zhang, T., Harrington, K. B., & Sherf, E. N. (2022). The errors of experts: When expertise hinders effective provision and seeking of advice and feedback. *Current Opinion in Psychology*, 43, 91–95. <https://doi.org/10.1016/j.copsyc.2021.06.011>
- Zou, W., Tang, L., Zhou, M., & Zhang, X. (2024). Self-disclosure and received social support among women experiencing infertility on reddit: A natural language processing approach. *Computers in Human Behavior*, 154, 1-11. <https://doi.org/10.1016/j.chb.2024.108159>

# Appendices

## Appendix A

### A.1 Epistemic-Aleatory Rating Scale (EARS)

---

#### Aleatoriness Subscale

---

A1 ... is something that has an element of randomness

A3 ... feels like it is determined by chance factors

---

#### Epistemicness Subscale

---

E3 ... is knowable in advance, given enough information

E4 ... is something that well-informed people would agree on

---

*Note.* Adapted from “Investor Behavior Under Epistemic vs. Aleatory Uncertainty” by D.J. Walters, G. Ülkümen, D. Tannenbaum, C. Erner, and C. R. Fox. 2023, *Management Science*, 69(5), p. 2765. (<https://doi.org/10.1287/mnsc.2022.4489>). Copyright 2023, Authors.

### A.2 Adviser Competence Rating Scale

---

How knowledgeable is this advisor?

How competent is this advisor?

How credible is this advisor?

How much do you trust this advisor?

Would you seek additional information or advice from this person in the future?

How persuasive is this advice?

How accurate is this advice?

---

*Note.* Reprinted from “Do people inherently dislike uncertain advice?” by C. Gaertig, and J. P. Simmons. 2018, *Psychological Science*, 29(4), p. 507. (<https://doi.org/10.1177/0956797617739369>). Copyright 2018, Authors.