Exploring Social Metacognition: the role of confidence in updating estimates of advisor reliability with and without feedback.

Matt Jaquiery

Wolfson College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2020

Abstract

This *R Markdown* template is for writing an Oxford University thesis. The template is built using Yihui Xie's `bookdown` package, with heavy inspiration from Chester Ismay's `thesisdown`, and the `OxThesis` LaTeX template (most recently adapted by John McManigle).

This template's sample content include illustrations of how to do the various things you need to write a thesis in R Markdown, and largely follow the structure from this R Markdown workshop.

Congratulations for taking a step further into the lands of open, reproducible science, by writing your thesis using a tool that allows you to transparently include tables and dynamically generated plots directly from the underlying data. Hip hooray!

Exploring Social Metacognition: the role of confidence in updating estimates of advisor reliability with and without feedback.

Matt Jaquiery

Wolfson College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2020

For TBC

Acknowledgements

This is where you will normally thank your advisor, colleagues, family and friends, as well as funding and institutional support. In our case, we will give our praises to the people who developed the ideas and tools that allow us to push open science a little step forward by writing plain-text, transparent, and reproducible theses in R Markdown.

We must be grateful to John Gruber for inventing the original version of Markdown, to John MacFarlane for creating Pandoc (`http://pandoc.org`) which converts Markdown to a large number of output formats, and to Yihui Xie for creating `knitr` which introduced R Markdown as a way of embedding code in Markdown documents, and `bookdown` which added tools for technical and longer-form writing.

Special thanks to Chester Ismay, who created the `thesisdown` package that helped many a PhD student write their theses in R Markdown. And a very special tahnks to John McManigle, whose adaption of Sam Evans' adaptation of Keith Gillow's original maths template for writing an Oxford University DPhil thesis in LaTeX provided the template that I adapted for R Markdown.

Finally, profuse thanks to JJ Allaire, the founder and CEO of RStudio, and Hadley Wickham, the mastermind of the tidyverse without whom we'd all just given up and done data science in Python instead. Thanks for making data science easier, more accessible, and more fun for us all.

Ulrik Lyngs
Linacre College, Oxford
2 December 2018

Abstract

This *R Markdown* template is for writing an Oxford University thesis. The template is built using Yihui Xie's `bookdown` package, with heavy inspiration from Chester Ismay's `thesisdown`, and the `OxThesis` LaTeX template (most recently adapted by John McManigle).

This template's sample content include illustrations of how to do the various things you need to write a thesis in R Markdown, and largely follow the structure from this R Markdown workshop.

Congratulations for taking a step further into the lands of open, reproducible science, by writing your thesis using a tool that allows you to transparently include tables and dynamically generated plots directly from the underlying data. Hip hooray!

Contents

List of Figures

List of Tables

List of Abbreviations

1-D, 2-D . . . One- or two-dimensional, referring in this thesis to spatial dimensions in an image.

Otter . . . . . One of the finest of water mammals.

Hedgehog . . . Quite a nice prickly friend.

```
knitr::opts_chunk$set(echo = F)
```

Part I

Introduction

1 Introduction

This thesis attempts to determine whether individual psychological processes in the seeking and taking of advice are sufficient to produce entrenched biases at the network level. The work includes empirical investigation into the individual psychological processes underlying advice seeking and advice taking using behavioural experiments; computational modelling and behavioural experimentation exploring the effects of contextual factors on advice-taking; agent-based computational modelling of the effects interactions between agents with psychological advisor evaluation processes that produce biases in assimilation of information and source selection; and a comparison between the structures of networks produced in these models and those naturally occurring social networks. The organisation is as follows: this introduction establishes the core concepts invoked in this thesis, and describes their treatment in the literature; the following sections each include a short chapter on the specific question addressed and the techniques used, a detailed description of the work conducted, and a short discussion of the conclusions drawn from the work; and a final section offers broader conclusions arising as a consequence of the presented work, alongside some suggestions for related research.

Overview

The first line of behavioural experiments report investigates whether people decide which advisors to consult using internal signals where external feedback is unavailable, extending a previous finding concerning the influence of advice (Pescetelli and Yeung 2018) into the domain of advisor choice. While little support is found for the metacognitive mechanisms underpinning the influence of advice, the frequency of agreement is shown to be a good indicator of advisor choice. Next, the effects of advisor agreement are explored in terms of influence using a different decision-making domain; a date estimation rather than perceptual decision task. Again, people denied objective feedback on the quality of advice are shown to attend to agreement.

The work above demonstrates how the influence of advice depends upon the reputation of an advisor built up over time. In a second line of experiments, reputations of advisors are established at the outset and the responses to different pieces of advice are explored. People consider both the properties of advice and the properties of advisors, placing less trust in advice which appears suspicious or which comes from a suspicious source. These behavioural results support implications from evolutionary models which demonstrate that the presence of even a few bad actors in a population can mean that distrusting all advice from all sources a little is adaptive. The evolutionary models are extended to show that this kind of general distrust, known as egocentric discounting, emerges as adaptive even where none of the agents deliberately mislead others, and even where all agents are equally skilled and fully cooperative.

Finally, network effects arising from interconnected networks of agents are explored. Agents are based on the empirical results heretofore presented, and the consequences are explored for networks with various starting structures characterised by sparsity and homogeneity, under varying rates and reliabilities of external feedback. !TODO[results of this]. The resulting network characteristics are compared with network structures from real online networks from social media websites. !TODO[results of this].

Advice

Advice is broadly defined as information which comes from a social source. Advice is therefore different from other sources of information in that it is the result of (at least) mental processing of other information. In some cases, it may additionally include discussions among different group members (e.g. advice from the International Advisory Panel on Climate Change). Throughout this thesis, the focus is primarily on advice which comes from a single, stable source, as when we see a post by an acquaintance on social media, or when a stranger provides us with advice.

Advice-taking

Advice occurs in the context of a decision, and forms a part of the information which is integrated during the decision-making process to produce a decision. To the extent that the decision reached differs from the decision that would have occurred had the advice not been presented, the advice has had an effect on the decision; to the extent that this difference changes the decision in a way consistent with the advice, the advice has been 'taken' (as opposed to 'rejected').

It is tacitly implied by many operationalizations of advice-taking that the informational content of the advice determines the extent to which it is taken or rejected `citation needed`. Insofar as the identity of the advisor matters, it matters because it functions as a cue to the informational content of the advice. This is likely a major oversimplification, however, because in many real-world contexts advice-giving and advice-taking form part of a developing social relationship: being consulted for advice and having one's advice followed are inherently rewarding (Hertz and Bahrami 2018; Hertz, Palminteri, et al. 2017); and taking advice can serve as a (sometimes costly) social signal of valuing a relationship with a person or group (Byrne et al. 2016). Furthermore, some authors have argued that people may perceive taking advice as sacrificing their independence or autonomy (**CITATIONNEEDED**)!TODO. While this thesis follows previous literature in omitting to consider the wider social concerns influencing the taking of advice, it is nevertheless important to remember that the processes investigated herein take place in a variety of social contexts where complex social agents attempt to optimise over numerous goals over numerous timescales.

Three-factor model of trust

The degree to which advice is taken is proportional to the trust placed in the advisor by the decision-maker. Interpersonal trust, or the degree to which one is prepared to place one's fortune in the hands of another (e.g. by relying on their advice), is apportioned by Mayer et al. (1995) onto three properties of the advisor (as judged by the decision-maker): ability, benevolence, and integrity. To these three properties of the advisor we may add the decision-maker's general propensity to trust, as well as situational cues and task cues (e.g. the phenomenon that advice is more readily taken for hard tasks than easy ones, (Gino and Moore 2007)).

Ability

Ability captures the expertise of an advisor: their raw ability to perform the task for which they are giving advice. In some cases this is relatively straightforward, as in the expertise of a General Practitioner in matters of health and disease, and in others more complex, as in the expertise of a hairdresser when deciding on a haircut (when matters of personal taste comingle with aesthetic considerations of facial structure, practical considerations of hair constitution, and social considerations of fashion). The greater the ability of an advisor, the greater the influence of their advice, as demonstrated by experiments showing that participants' decisions are more affected by the advice of advisors who are labelled as more expert in a relevant domain (Sah et al. 2013; Schultze et al. 2017; Sniezek, Schrah, et al. 2004; Sniezek and Van Swol 2001; Soll and Mannes 2011), or are shown to be more expert empirically (Pescetelli 2017; Sah et al. 2013; Ilan Yaniv and Kleinberger 2000).

Benevolence

Benevolence refers to the extent to which the advisor seeks to further the interests of the decision-maker. Where ability represents the absolute limit on the quality of advice, benevolence represents the extent to which the advice approaches this limit. The advice of even a renouned expert may be doubted if there is reason to believe their goal is to mislead, a vital lesson for medieval monarchs with their councils of politicking advisors. Experimental work has shown that psychology students relied more on the advice of their friends than on the advice of labelled experts (**CITATIONNEEDED**)!TODO, and that participants are more inclined to reject advice when uncertainty is attributed to malice rather than ignorance (Schul and Peni 2015).

Integrity

Advisors with integrity exhibit adherence to principles which the decision-maker endorses. As with benevolence, integrity acts to determine the extent to which advice approaches the limit imposed by ability. While not mutually exclusive, integrity is typically important where relationships are less personal (e.g. we may place great trust in a General Practitioner because of their expertise in medical matters and their *integrity* in adhering to a set of professional ethical and conduct requirements). `Some description of the research`(**CITATIONNEEDED**)!TODO

Normative models of advice-taking

Advice-taking can be evaluated formally with reference to a normative model. The simplest and most common of these views the decision-making task as an estimation problem (or combination of estimation problems), and provides an approximately Bayesian variance-weighted integration of independent estimates. To borrow from Galton (1907), consider the task of judging the weight of a bullock. We can model any single guess (e) as the true weight (v) plus some error (ϵ):

$$e = v + \epsilon \tag{1.1}$$

The key insight is to observe that the error is drawn from a normal distribution ($\mathcal{N}(\mu = 0, \sigma^2)$)[1]. As the number of samples from this distribution increases, the mean of those samples tends towards the mean of the distribution. Thus, the more estimates are taken, the closer on average the sum of errors will be to 0.

$$\frac{\sum_i^N(e_i)}{N} = \frac{\sum_i^N(v + \mathcal{N}(\mu = 0, \sigma_i^2))}{N} \tag{1.2}$$
$$\frac{\sum_i^N(e_i)}{N} = \frac{\sum_i^N(v)}{N} + \frac{\sum_i^N(\mathcal{N}(\mu = 0, \sigma_i^2))}{N} \tag{1.3}$$
$$\frac{\sum_i^N(e_i)}{N} = \frac{Nv}{N} + \hat{0} \tag{1.4}$$
$$\frac{\sum_i^N(e_i)}{N} \approx v \tag{1.5}$$

Observe that this formulation is true no matter the value of N. On average, it is always better to have more estimates than fewer. This suggests that, even in the situation where there are only two estimates (the decision-maker's and the advisor's), the best policy will be to incorporate both estimates into the final decision.

The variance of the normal distribution from which errors are derived (σ_i^2) is, in the example above, drawn from a normal distribution itself ($\sigma_i^2 \sim \mathcal{N}(\mu = 0, \mathrm{sd}^2)$meaning that it is also cancelled out on average over repeated samples). Where few estimates are taken, weighting those estimates by the variance of the error distributions will increase the accuracy of the estimates in proportion to the difference between the variances. Many experimental implementations of this model avoid weighting issues by calibrating decision-makers and advisors to be equally accurate on average ($\sigma^2_{\text{decision-maker}} = \sigma^2_{\text{advisor}}$). The result of this constraint is that the optimal policy is simply to average all estimates together.

[1] The normal distribution is well-supported by empirical evidence, but note that any symmetrical distribution around 0 will lead to the same conclusion.

Egocentric-discounting

From the perspective of the normative model above, decision-makers should weigh their own estimate equally with each other estimate they receive in the process of coming to their decision. One of the most robust findings in the literature on advice-taking is that people routinely underweight advisory estimates relative to their own estimates, a phenomenon known as *egocentric discounting*(Dana and Cain 2015; Gino and Moore 2007; Hütter and Ache 2016; Liberman et al. 2012; Minson and Mueller 2012; Rader et al. 2017; Ronayne and Sgroi 2018; See et al. 2011; Soll and Mannes 2011; Trouche et al. 2018; Ilan Yaniv and Kleinberger 2000; Ilan Yaniv and Choshen-Hillel 2012; Ilan Yaniv and Milyavsky 2007). Egocentric discounting occurs in both feedback and no-feedback contexts (Ilan Yaniv and Kleinberger 2000).

Explanations for egocentric discounting are usually framed in terms of personal-level psychology: decision-makers have better access to reasons for their decision (Ilan Yaniv and Kleinberger 2000); overrate their own competence (Sniezek, Schrah, et al. 2004); may have a desire to appear consistent (Ilan Yaniv and Milyavsky 2007); may see opinions as possessions (Soll and Mannes 2011); may be loss-averse to providing a worse final estimate due to advice-taking (Soll and Mannes 2011); or have difficulty avoiding anchoring (Schultze et al. 2017) or repetition bias effects (Trouche et al. 2018). None of these explanations has survived rigorous empirical testing, however, and recently suggestions have widened to include consideration of aggregate-level rather than personal-level causes, with Trouche et al. (2018) arguing that the potential for misaligned incentives between decision-maker and advisor motivate discounting of advice.

In the course of this thesis `crossref needed`, I demonstrate that egocentric discounting may be a stable metastrategy which protects against exploitation, carelessness, incompetence, and miscommunication. From this perspective, the normative model pertains to a particular instantiation of a problem with questionable ecological validity given the typical ethology of advice-taking in humans. While such considerations affect the conclusions one draws from egocentric discounting relative to the normative model, they do not detract substantially from the practice of using the normative model as an optimum 'set point' from which to evaluate advice-taking behaviour.

Evaluation of advice

The value of advice, in informational terms, is measured in relation to the optimal decision. It is necessary to distinguish two perspectives on the value of advice: the optimality of the decision recommended by the advice itself; and the optimality of the decision based on advice relative to the decision which would have been made had the advice not been received. This distinction is necessary because people alter their advice-giving behaviour in anticipation of discounting on the part of the decision-maker (**CITATIONNEEDED**; for a case in human-machine teaming see Azaria et al. 2016)!TODO, somewhat akin to starting negotiations with a higher demand than one is hoping to settle for.

A single piece of advice can be evaluated using its own properties and the properties of the advisor giving the advice. Furthermore, that evaluation can serve to update the properties of the advisor. A piece of advice's own properties will include its plausibility (e.g. participants in estimation tasks discount advice which is distant from their own initial estimates more heavily (I. Yaniv 2004)), while the properties of the advisor will include the advisor's trustworthiness (see above). The updating of trust following experience of advice is likely to be largely in the domain of ability, although other domains may be affected where the advice is particularly egregious.

Updating advisor evaluations

While a single piece of advice must be taken on its own terms, people can construct relatively accurate estimates of advisors' advice when provided with feedback on the decisions they use the advice to make (Pescetelli 2017; Sah et al. 2013; Ilan Yaniv and Kleinberger 2000). This likely happens as an analogue of reinforcement learning, where feedback allows an error signal to be used to update the estimate of the advisor's ability ($\hat{s}^a$) rather than one's own beliefs about the world, according to some learning rate (λ).

$$\text{this is wrong; need to check RL models for an analogue} \tag{1.6}$$
$$\hat{s}_{t+1} = \hat{s}_t + |e^a_t - v| \cdot \lambda \tag{1.7}$$

Advisor evaluation without feedback

Where feedback is not available, participants in experiments continue to demonstrate an ability to respond rationally to differences in advisor quality (Pescetelli 2017). This is evidently not done through access to the correct real-world values, because feedback providing those values is unavailable, and, were participants aware of those values themselves, it stands to reason they would have provided those values (and thus not require advice!). Pescetelli and Yeung (2017) suggest the mechanism for this ability to discriminate between advisors in the absence of feedback is performing updates based on confidence-weighted agreement.

Agreement

Consider first the non-weighted agreement case, where the advisor's estimate at time t (e^a_t) and the decision-maker's estimates (e^d) are binary ($\in 0, 1$). The estimate of the advisor's ability ($\hat{s}^a$) is updated positively if the advisor and decision-maker agree, and negatively otherwise, according to the learning rate λ.

$$\hat{s}^a_{t+1} = \hat{s}^a_t + (-2\ |e^d_t - e^a_t| + 1) \cdot \lambda \tag{1.8}$$

Confidence-weighted agreement

The updating of advice contingent on agreement may be weighted by confidence in the initial decision (c^d), such that agreement and disagreement are considered more informative about the quality of the advice when the decision with which they agree or disagree is more certain.

$$\hat{s}^a_{t+1} = \hat{s}^a_t + \left(-2 \left|e^d_t - e^a_t\right| + 1\right) \cdot c^d_t \cdot \lambda \tag{1.9}$$

Continuous estimate case

Homophily and echo-chambers

Homophily is the ubiquitous phenomenon that individuals more closely connected to one another within a social network tend to be more similar to one another than would be expected by chance across numerous dimensions, from demographics to attitudes (McPherson et al. 2001). Whether homophily in virtual social networks is responsible for increases in polarisation is debated. Proponents point to increases in polarisation (e.g. in politics: `Pew Research Center, 2014`!TODO), to empirical studies demonstrating homophily in virtual social networks (Cardoso et al. 2017; Colleoni et al. 2014), and to studies examining selective exposure online (Kobayashi and Ikeda 2009), and to echo chambers: egregious examples of highly homophilous networks with pathological polarisation (Sunstein 2002; Sunstein 2018). The empirical components of the argument are contested, with evidence that virtual social networks are less homogenous than offline social networks (and hence depolarising, Barberá 2015), and that selective exposure is a somewhat dubious finding which does not show up clearly online (Garrett 2009a; Garrett 2009b; Nelson and Webster 2017; Sears and Freedman 1967). Modelling work demonstrates, however, that where there is a bias in assimilation of information, homophily exacerbates polarisation (Dandekar et al. 2013). Where polarisation in turn increases homophily, for example through selective exposure or avoidance, a self-reinforcing spiral emerges wherein social connections become increasingly homogenous and attitudes increasingly extreme (H. Song and Boomgaarden 2017).

Source selection and information weighting

Context-dependency of epistemic processes

Part II

Psychology of advice

Contents

2

Psychological mechanisms of advisor evaluation

Where feedback is unavailable, people may use their own sense of certainty as a yardstick for evaluating advice (Pescetelli and Yeung 2018; Pescetelli 2017): advisors who agree when one is confident are perceived as more helpful; while those who disagree when one is confident are perceived as less helpful. Confidence serves as a proxy for objective feedback, and functions well in this role insofar as the judge has high metacognitive resolution (i.e. higher confidence is indicative of a greater probability of being correct).

Judge-advisor system

Table of experiments

Advisors	Choice	Task	Feedback	Result
AiC vs AiU	Yes	Perceptual, binary (MATLAB)	No	Suggestive
AiC vs AiU	Yes	Perceptual, binary	No	Inconclusive

Advisors	Choice	Task	Feedback	Result
In/accurate	Yes	Perceptual, binary	No	Accuracy selected more often
Low/High agreement	Yes	Perceptual, binary	No	Agreement selected more often
Accurate vs Agreement	No	Estimation, continuous	Yes and No	Accurate preferred given feedback, agreement preferred without feedback

3
Psychology of advice-taking

3.1 Playground

!TODO[remove playground]

4
Psychology of source selection

4.1 Experiment 1: Metacognitive-contingent advisors (Lab)

Pescetelli et al. (2018) showed that, in the absence of objective feedback, advice was more influential coming from an advisor who agrees with a participant when that participant is confident (*Bias Sharing*) than coming from an advisor who agrees with a particpant when that participant is unconfident (*Anti Bias*). This provides evidence of a metacognitive sensitivity in the tracking of advice and the updating of advisor utility. Here we investigate whether these effects show up in the domain of advisor selection.

The literature on information exposure and evaluation indicates that people evaluate more favourably information which agrees with their currently-held opinion !TODO[REF], and preferentially seek out information sources which are likely to provide information which agrees with their currently-held opinion (Garrett 2009a; Sears and Freedman 1967). If this holds true in the context of the judge-advisor system, advice from *Bias Sharing* advisors ought to be evaluated more favourably (influence should increase) and should be sought more frequently. Given the evidence in favour of the first of these propositions, we here investigate the latter: given a choice, will judges prefer to receive advice from a *Bias Sharing* advisor over receiving advice from an advisor who does not share the judge's bias?

Pescetelli et al. (2018) used a judge-advisor system to demonstrate that judges are influenced to a greater extent by advisors who share their biases. Participants played the role of judge in a judge-advisor system, while the advisors were virtual agents whose advice-giving was dependent upon the confidence and correctness of the judges' initial decisions. The advisors were balanced for overall agreement with the judge and objective correctness of advice. We place participants in a similar paradigm in which they are given a choice between advisors, and hypothesise that they will more frequently seek advice from the *Bias Sharing* advisor than from the *Anti Bias* advisor.

4.1.1 Open science

Preregistration

The study was preregistered using AsPredicted.org. The preregistration document can be found at `https://aspredicted.org/ze3tn.pdf`.

Open materials

Experimental materials, including scripts required to run the experiments in MATLAB and scripts required to analyse the data in R, are available from the GitHub repository.

The experimental design was adapted from Pescetelli et al. (2017), and the major work in the design, as well as the experimental scripts, is due to Niccolo. The full list of changes to the final design can be seen in the commits to the project repository, which began as a fork of Niccolo's work.

Open data

Anonymised study data can be found at !TODO[Use a sensible archive format for this study data, archive on OSF, and produce data dictionary]. A preloaded version of the data formatted appropriately for the R scripts is included in the GitHub repository.

4.1.2 Method

Participants

24 participants ($M_{age} = 22 \pm SD\ 4.7$, 5 male, 19 female, 0 other) recruited from the University of Oxford participant recruitment platforms took part in the experiment. An additional 2 participants attended experimental sessions but their data were not analysed. Participants were compensated for their time with either course credit for a psychology degree, or 10GBP.

Ethics

Ethical approval for the study was granted by the University of Oxford Medical Sciences Interdivisional Research Ethics Committee (Ref: R55382/RE001).

Procedure

The experiment consisted of a judge-advisor system with a perceptual decision task (Figure 4.1). Participants played the role of the judge, and the advisors were played by virtual agents whose answers depended upon the confidence with which the judge reported the initial decision. In the majority of trials (92%), participants were offered advice from virtual advisors. In one third of these trials ('choice trials'), participants chose which advisor to receive advice from by clicking on their respective portraits appearing at the top and bottom of the screen. On the remaining two thirds of trials ('forced trials'), participants were forced to take advice from one of the two advisors (equiprobably). On these trials, the forced advisor's portrait appeared at the top or bottom of the screen, with a red cross appearing in the other location, which was not selectable. On the remaining 8% of trials, participants received no advice and were given no opportunity to revise their initial decision. These 'catch trials' were included to encourage participants to attend to the initial decisions.

Each participant completed 363 trials (51 practice trials over 2 blocks + 12 x 26-trial experimental blocks) in which they had to identify the box with the most dots (Figure 4.1). The difficulty of the task was continually adjusted throughout the experiment using a 2-down, 1-up staircase procedure to keep the participant's

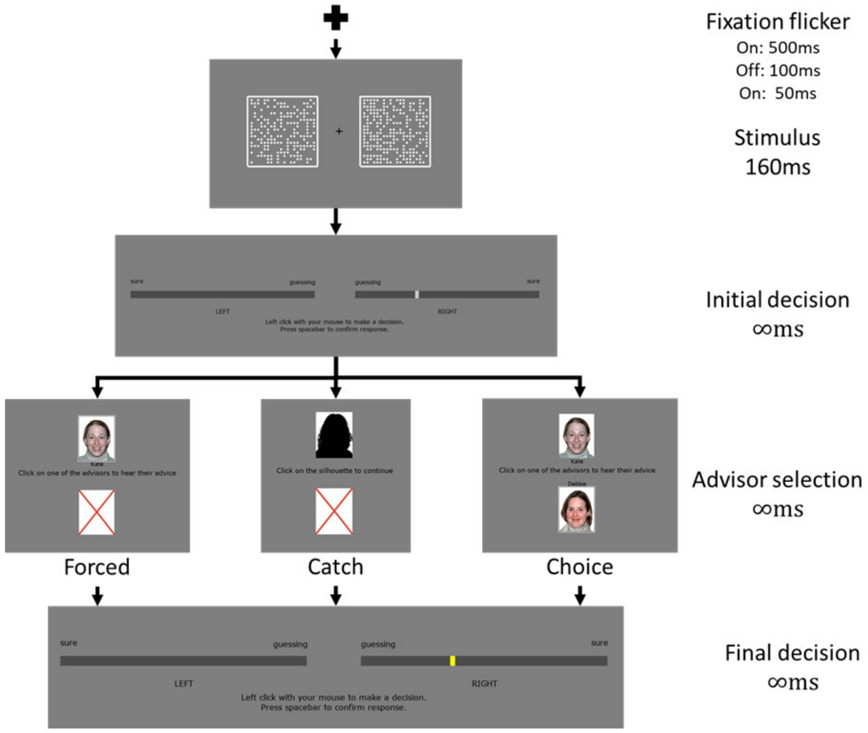

Figure 4.1: Experiment 1 procedure: The task began with a blank screen containing only a fixation cross and progress bar. Momentarily prior to the onset of the stimuli the fixation cross flickered. The stimuli, two rectangles containing approximately 200 dots each, appeared for 0.16s, one on either side of the fixation cross. Once the stimuli disappeared, a response-collection screen appeared and prompted the participant to indicate their initial decision and its confidence by selecting a point within one of two regions. The left region indicated a decision that the target was on the left, and increasingly-leftwards points within that region indicated increasing confidence in that decision. The right region indicated a decision that the target was on the right, and increasingly-rightwards points within that region indicated increasing confidence in that decision. Next, the participant was presented with a choice screen. The choice screen displayed two images, one at the top of the screen and one at the bottom. The images were one of the following: an advisor portrait, a silhouette, or a red cross. The red cross was not selectable, forcing participants to choose the other option. The silhouette offered no advice, and was only ever offered as a forced choice. Selecting an advisor image provided the participant with the opinion of that advisor on the trial. Having heard the advice, the participant was again presented with the response-collection screen, with a yellow indicator marking their original response. A second (final) judgement was collected using this screen (except on catch trials), and the trial concluded.

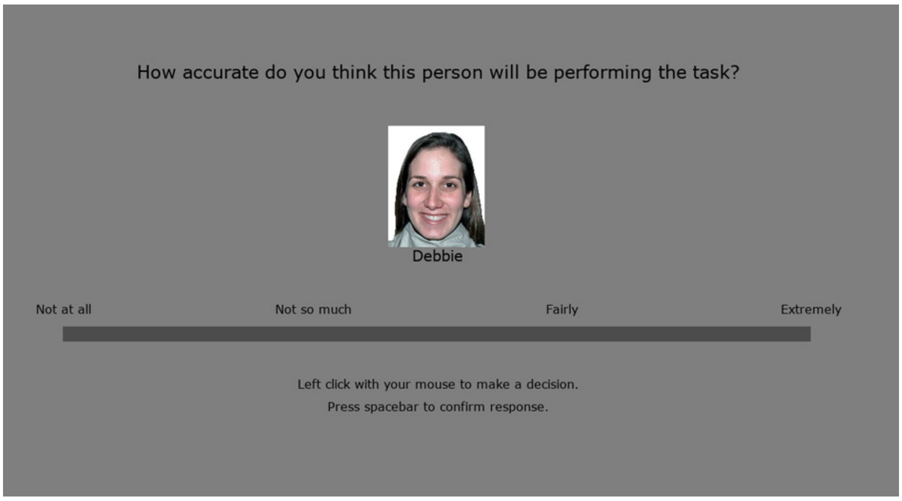

Figure 4.2: Experiment 1 advisor questionnaire: Participants rated advisors on a number of different dimensions.

initial decision accuracy at 72%. At the end of each block, participants were notified as to their (final decision) accuracy in the block and given the opportunity to rest for as long as they wished. Throughout the experiment a progress bar provided a graphical indication of the number of trials remaining in the experiment.

After each block participants were told what percentage of the (final) answers they had provided were correct and allowed to take a short, self-paced break. Prior to the first experimental block, after the final experimental block, and after the 4th and 8th experimental blocks, participants were presented with a questionnaire (Figure 4.2). The questionnaire contained 4 questions for each advisor. The questions asked for the judge's assessment of the advisor's likeability, trustworthiness, influence, and ability to do the task. The questions presented before the first experimental block were worded prospectively (e.g. 'How much are you going to like this person?' as opposed to 'How much do you like this person?'). Answers were provided by moving a sliding scale below the advisor's portrait towards the right for more favourable responses (marked 'extremely') or towards the left for less favourable responses (marked 'not at all').

Each participant attended the experiment individually, was welcomed and briefed on the experimental procedure, and had their informed consent recorded, before the experiment began. They were seated a comfortable distance in front of a 24'

Table 4.1: Table : Advisor advice profiles

	Initial decision confidence	Probability of agreement (%)	
		Bias Sharing	Anti Bias
Participant correct	High (top 30%)	80	60
	Medium (middle 40%)	70	70
	Low (bottom 30%)	60	80
Participant incorrect	Any	30	30
Total agreement	Participant correct	70	70
	Participant incorrect	30	30

(1440x900 resolution) computer screen in a small, quiet, and dimly-lit room. The experiment took place wholly on the computer, and lasted around 45 minutes.

The experiment was programmed in MATLAB R2017b (*MATLAB* 2017) using the Psychtoolbox-3 package (Kleiner et al. 2007).

Advisor advice profiles

The advisors are virtual agents whose probability of agreeing with the participant's decision varies as a function of the participant's confidence and correctness in the initial decision phase. Table 4.1 illustrates how this relationship functions, and shows that the overall correctness and agreement rates of the advisors is equivalent overall. Importantly, on largest minority of trials, the middle 40%, the advisors are exactly equivalent, meaning these trials can be compared directly without confounds arising from agreement rate and initial confidence.

Analysis

Data analysis was performed using R (R Core Team 2018). For a full list of packages and software environment information, see !TODO[figure out where to include this stuff. Appendix? Also link to a containerized version of this.]

Bayes Factors (BF) are presented alongside p values and test statistics. A BF $<$ 0.33 indicates decisive evidence in favour of the null hypothesis over the alternative hypothesis (with lower values being increasingly clear), BF $>$ 3 indicates decisive evidence of the alterative over the null (with higher values being increasingly clear), and 0.33 $<=$ BF $<=$ 3 indicates there is insufficient evidence to reach a conclusion.

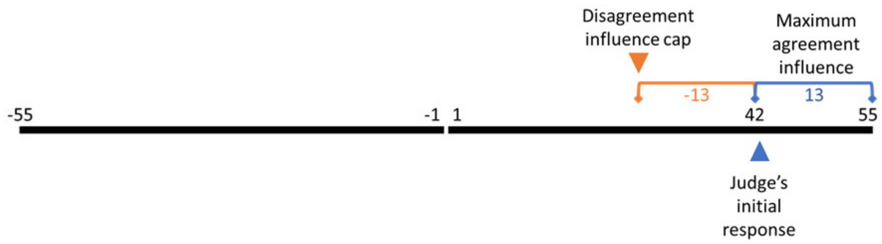

Figure 4.3: Capping influence to avoid scale bias: In this example the judge's initial response is 42, meaning that their final decision could be up to 13 points more confident or up to 97 points less confident. Any final decision which is more than 13 points less confident is therefore capped at 13 points less confident.

Influence, the dependant variable in some analyses, is calculated as the extent to which the judge's initial decision is revised in the direction of the advisor's advice. The initial (C_1) and final (C_2) decisions are made on a scale stretching from -55 to +55 with zero excluded, where values <0 indicate a 'left' decision and values >0 indicate a 'right' decision, and greater magnitudes indicate increased confidence. Influence (I) is given for agreement trials by the shift towards the advice:

$$I|\text{agree} = f(C_1) \begin{cases} C_2 - C_1 & C_1 > 0 \\ -C_2 + C_1 & C_1 < 0 \end{cases} \tag{4.1}$$

And by the inverse of this for disagreement trials:

$$I|\text{disagree} = -I|\text{agree} \tag{4.2}$$

The confidence scale excludes 0, and thus the final decision can always be more extreme when moving against the direction of the initial answer than when moving further in the direction of the initial answer. A capped measure of influence was used to minimise biases arising from the natural asymmetry of the scale. This measure was calculated by truncating absolute influence values which were greater than the maximum influence which could have obtained had the final decision been a maximal response in the direction of the initial answer (Figure 4.3).

The capped influence measure I_{capped} is obtained by:

$$I_{\text{capped}} = f(C_1) \begin{cases} \min(I, 2C_1 - 55) & C_1 > 0 \\ \max(I, 2C_1 + 55) & C_1 < 0 \end{cases} \tag{4.3}$$

The explicit measure of trust is obtained using questionnaires. The questionnaires are delivered at 4 time points, and consist of 4 questions per advisor which are answered on a 1-100 scale.

4.1.3 Result

4.1.4 Discussion

4.2 Experiment 2: Metacognitive-contingent advisors

4.2.1 Method

4.2.2 Result

4.2.3 Discussion

4.3 Experiment 3: Advisor accuracy

4.3.1 Method

4.3.2 Result

4.3.3 Discussion

4.4 Experiment 4: Advisor agreement

4.4.1 Method

4.4.2 Result

4.4.3 Discussion

4.5 Experiment 5: Accuracy vs agreement (Date estimation)

4.5.1 Method

4.5.2 Result

4.5.3 Discussion

4.6 General discussion

Part III

Context of advice

5 Context of Advice

Optimal advice policy differs by context.

6

Sensitivity of advice-taking to context

Advice-taking is often overly-conservative as compared to the normative level of advice-taking for a given experimental design. I argue that participants' performances in advice-taking experiments reflect both the specifics of the experimental design and prior expectations about advice-taking situations. These prior expectations may be both learned, as where individuals who grow up in less stable environments show lower propensity to trust `reference?`!TODO, and inherited. Most useful for the current argument would be a demonstration that conservatism can emerge within a population even where detrimental advice is rarely experienced, and that this can thus produce individuals who exhibit conservatism without ever experiencing detrimental advice. This demonstration is presented in the form of evolutionary modelling.

As discussed in the previous chapter, conservatism is optimal under some circumstances, and thus we expect that simulated agents allowed to evolve an advice-taking policy in those circumstances will evolve a conservative policy. I explored this tendency as a function of three plausible scenarios. The first scenario is one in which agents occasionally give deliberately poor advice to their advisee, which represents situations where advisors' interests may sometimes be contrary to judges' interests, unbeknownst to the judges. In the second scenario, advice is simply noisier than the judge's own initial estimate, either because the judge is less competent at the task, less willing to exert the required effort for the task, or because the advice is communicated imperfectly. In the third scenario, agents belong to either a 'cautious' or a 'confident' group in how they express and interpret advice, which is a simple analogue of the observation that people's expressions of confidence are idiosyncratic (Ais et al. 2016; Navajas et al. 2017). In each of these three scenarios, it is hypothesised that some level of egocentric discounting will emerge as the dominant strategy, i.e., the mean population weighting for initial estimates versus advice will be greater than .50.

6.1 General method

Agent-based computational models of an evolutionary process were programmed in R (R Core Team 2018) and run variously on a home computer and the Oxford Advanced Research Computing cluster (Richards 2015). The code is available at `https://github.com/oxacclab/EvoEgoBias`, and the specific data presented below are archived at !TODO.

The models reported here use 1000 generations of 1000 agents which each make 30 decisions/generation on which they receive the advice of another agent. Decisions are either point estimation (Scenarios 1 and 2) or categorical decision with confidence (Scenario 3). Each agent combines their own initial estimate with the advice of another agent, with the relative weights of the initial estimate and advice set by the agent's egocentric bias parameter, to produce a final decision. Final decisions are evaluated by comparison with the objective answer, and an agent's fitness is the sum of its performance over the 30 decisions of its lifetime.

6.1.1 Initial estimates

The agents perform a value estimation (category estimation in Scenario 3) task. Agent i's initial estimate t is the true value (v_t), plus some noise drawn from a normal distribution with mean 0 and standard deviation equal to the agent's insensitivity parameter (s^i, which is itself drawn from a positive-clamped normal distribution with mean and standard deviation 10 when the agent is created).

An agent's initial estimate (e^i_t) is thus:

$$e^i_t = v_t + N(0, s^i) \tag{6.1}$$

6.1.2 Advice

Each agent receives advice from another agent which it combines with its initial estimate to reach a final decision. The advice has a probability of being mutated in some fashion. The mutation depends upon the scenario and is described separately for each.

6.1.3 Final decisions

In the basic model from which other models inherit their decision procedure, agent i produces a final decision t as the average of the agent's initial estimate (e^i_t) and another agent's advice (a^i_t), weighted by the agent's egocentric bias (b^i). The models typically change the value of a^i_t, which is typically a function of some other agent j's initial estimate e^j_t.

An agent's final decision (d^i_t) is thus:

$$d^i_t = \frac{e^i_t \cdot b^i + a^i_t \cdot (1 - b^i)}{2} \tag{6.2}$$

The final decisions in Scenario 3 are more complex, but follow a similar structure.

6.1.4 Reproduction

Roulette wheel selection is used to bias reproduction in favour of agents performing best on the decisions. Performance is determined by a fitness function which differs slightly between categorical and continuous decisions. For scenarios 1 and 2, which use continuous decisions, this fitness is obtained by subtracting the absolute difference between the final decision and the true value for each decision:

$$u^i = -\sum_{t=1}^{30} |v_t - d_t^i| \tag{6.3}$$

The selection algorithm proceeds as follows: The worst performance is subtracted from each agent's fitness and 1 added to put fitness scores in a positive range. These scores are then continually multiplied by 10 until the lowest score is at least 10 to improve resolution. Each agent is then given a probability to reproduce equal to their share of the sum of all fitness scores:

$$r^i = \frac{u^i}{\sum_{j=1}^{n} u^j} \tag{6.4}$$

where n is the number of agents and u has undergone the transformations described above.

Reproducing agents pass on their egocentric bias to their offspring. Other agent features, e.g. decision-making accuracy, are randomised when they are created. In the present simulations, agents receive no feedback on decisions, and cannot learn about or discriminate between their advisors. The key outcome of interest in each simulation is whether the population evolves towards egocentric discounting as the dominant adaptive strategy.

6.2 Scenario 1: misleading advice

In scenario 1, agents sometimes choose to offer misleading advice to their advisee.

6.2.1 Method

The true value (v_t) is fixed at 50 in this scenario. The agents do not learn about the true value over time, so a fixed and arbitrary value does not alter the results of the simulation.

Advice in this scenario is either the advising agent's initial estimate (e^j_t), or an extreme answer in the opposite direction to the advising agent's initial estimate (i.e. lower than 50 if e^j_t was above 50, and vice-versa). !TODO[describe what we actually do with this work; it has moved on a bit since transfer report]

6.2.2 Results

Where advice is always genuine, egocentric discounting does not emerge. Where advice is sometimes misleading, however, egocentric discounting emerges rapidly and remains stable throughout the rest of the stimulation.

6.2.3 Discussion

As one might expect, where there is the potential for exploiting trust it is safer to trust less, even if this reduces some of the benefits which would be gained from trusting. Egocentric discounting may be a viable strategy, even if there is no difference in ability between advice-seekers and advice-givers, given social contexts where the interests of advice-seekers and advice-givers are not perfectly aligned.

6.3 Scenario 2: noisy advice

In this scenario, agents offer advice which is of slightly lower quality on average than their own intial decisions. This could reflect a difference in effort or in expertise.

6.3.1 Method

As with Scenario 1, the true value (v_t) is fixed at 50.

Advice in this scenario has additional noise added:

$$a^i_t = e^j_t + N(0, x) \tag{6.5}$$

Where $x = 0$ in the no-noise condition and 5 in the noise condition.

Models were also run where a new decision was used as the basis for advice, rather than an initial estimate that some other agent would be using as a component in its own final decision:

$$a_t^i = v_t + N(0, s^j + x) \tag{6.6}$$

but the results were the same.

6.3.2 Results

As before, egocentric discounting emerges rapidly in the active condition and remains stable throughout the rest of the simulation.

6.3.3 Discussion

Where the advice is of worse quality on average, encoded here as additional variation, egocentric discounting tailors the relative weights of the estimates according to their average quality. As with situations in which an advice-seeker is more competent at making the relevant decision than their advisor, some measure of egocentric discounting is warranted in this scenario. It is worth noting that different competencies are not the only reason why advice may be systematically less valuable than initial decisions: difficulties in communicating the advice or different levels of conscientiousness in decision-making may also produce this effect.

6.4 Scenario 3: confidence confusion

While lackadaisical, incompetent, deliberately bad, or poorly communicated advice produces an obvious adaptive advantage for egocentric discounting, it is plausible that scenarios may exist where equally competent, wholly well-intentioned advice may still favour egocentric discounting. A common feature of advice is the communication of confidence, and this improves outcomes (Bahrami et al. 2010).

Notably, there are large and consistent individual differences in people's expressions of confidence (Ais et al. 2016; C. Song et al. 2011).

In this scenario agents are equally competent at the task, and do their best to assist one another, but may be hampered by expressing their estimates and advice with different confidences.

6.4.1 Method

The true value was drawn from a normal distribution around 50:

$$v_t = N(50, 1) \tag{6.7}$$

This allowed categorical answers which identified whether or not v_t was greater than 50. Agents were equiprobably assigned a personal confidence factor (c) of 1 or 10. This was used to scale the difference between the advising agent's intial estimate and the category boundary to produce the advice:

$$a_t^i = (e_t^j - 50 \cdot c^j) + 50 \tag{6.8}$$

Each agent then used the reciprocal of this process to translate advice back into its own confidence scale before integrating it with the initial estimate and arriving at a final decision:

$$d_t^i = \frac{e_t^i \cdot b^i + (\frac{1}{c^i}(a_t^i - 50) + 50) \cdot (1 - b^i)}{2} \tag{6.9}$$

This process amounts to the outgoing advice being translated into the advising agent's confidence language, and incoming advice being translated into the advised agent's confidence language. Where these languages are the same, the resulting advice is understood equivalently by both agents, but where there are differences the advice will be of greater or lesser magnitude compared to the initial estimate.

6.4.2 Results

Egocentric discounting again emerges in the condition where agents can have different confidence factors. The adaptiveness of egocentric discounting in this scenario arises because advice from a mismatched partner (e.g. one using a confidence scaling of 10 when the judge uses 1, or vice-versa) requires a different aggregation approach than advice from a matched partner. The application of an inappropriate aggregation approach results in misleading advice, so there is a pressure for a middle-of-the-road strategy whereby advice is counted, but not too much.

6.4.3 Discussion

Even where there is no intent to mislead, and no difference in basic ability, it is possible for egocentric discounting to emerge as an optimal strategy, purely because of differences in how agents communicate and understand estimates.

6.5 General discussion

The computational models show that egocentric discounting is an adaptive strategy in an array of plausible advice contexts: misleading advice, noisy advice, and different interpretations of confidence. While these models are necessarily limited in applicability to real life, they do demonstrate that egocentric discounting, while irrational for simple estimation problems with an objective answer and with advice that is not systematically better or worse than an individual's own judgement, may be beneficial for many of the kinds of decision for which we have sought and used advice in everyday life throughout our evolution.

This argument invites attention to the advice-taking task as much as to the properties of the advisor: it predicts that egocentric discounting will be attenuated where the outcome of decisions affects judges as well as advisors (Gino 2008, observed this effect but attributed it to judges falling prey to the sunk costs fallacy); where decisions rely more on objective than on subjective criteria (Van Swol 2011); where advisors and judges have opportunities to calibrate their confidence judgements with one another by completing training trials where they have to produce a shared decision with a shared confidence; and where incentives for judges and advisors are more closely aligned (Gino, Brooks, et al. 2012; Sniezek, Schrah, et al. 2004).

Notably, the utility of these heuristics does not depend on malice, mistake, or miscommunication: inconsistency in the usage of confidence terminology can produce adaptive pressure for egocentric discounting. More generally, the results indicate that properties of the advice-giving milieu can influence advice-taking strategies.

These models establish that it is plausible that people have deeply ingrained hyperpriors towards discounting advice. It is also possible, however, that people can flexibly respond to contexts, modulating their advice-taking appropriately. The voluminous experiments in advice taking discussed in the introduction to this section !TODO[that discussion?] can be seen as eliciting exactly this behaviour. In the next chapter I present new behavioural experiments which suggest that people do indeed modulate their advice-taking to the specifics of the context in which they are in.

7

Behavioural responses to advice contexts

7.1 Benevolence of the advisor population

The evolutionary models discussed in the last chapter demonstrated that optimal advice-taking strategies depend in part upon the advice one receives being a genuine effort to help. Difference in benevolence, or the extent to which the interests of the advisor and the judge overlap, have been shown to affect levels of advice-taking. !TODO showed that, where advisors were paid contingent upon the quality of the final decision, advice-taking was higher than where advisors were paid a flat fee for providing advice. !TODO[lit review stuff - inc. Behrens/Hunt? social decision making]

Advice-taking can be contingent on the properties of the advice, or on the properties of the advisor. In order to maximise the value of advice while minimising the potential exposure to exploitation, advice-taking should be contingent on a combination of these factors. Where advice is plausible it should be weighted relatively equally, whether it comes from an advisor who is sometimes misleading or not, but where advice is more implausible it should only be trusted when it comes from an advisor who is never misleading.

To explore whether people's behaviour matches this pattern, participants were recruited for a series of behavioural experiments in which they were given advice on

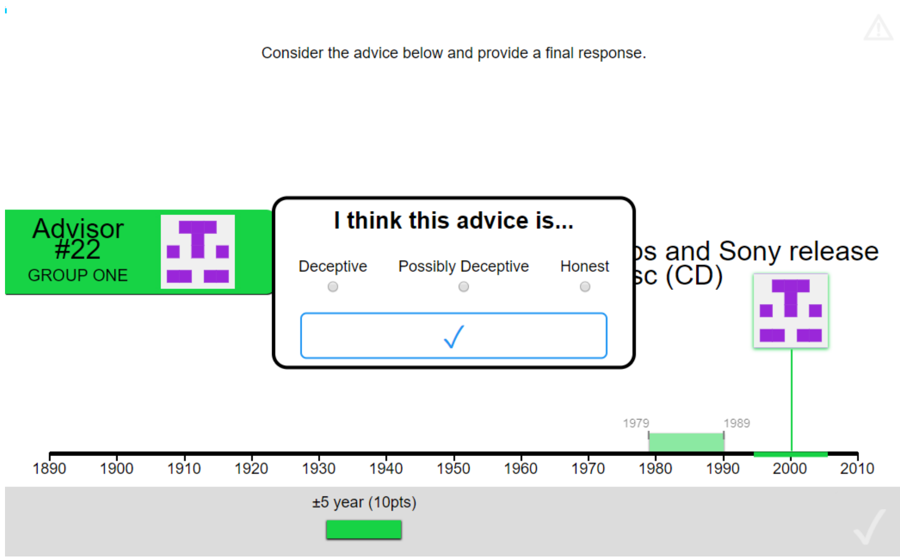

Figure 7.1: Advice honesty rating: participants rated advice on a three-point scale according to whether they thought the advice was deceptive or honest.

a date estimation task from advisors who were described as either always helpful or occasionally misleading. Results indicated that participants' advice-taking depended upon both the plausibility of the advice and the benevolence of the advisor, although the specific pattern differed from the one described above.

7.1.1 Method

Procedure

The procedure for these experiments follows the dates estimation task, described previously. Additionally, in this version, participants had to provide an assessment of the honesty of each advisory estimate they received before they could integrate it into their final decision.

Manipulation

Participants performed two blocks of experimental trials. In one block, participants received advice from an advisor who they were told would 'always try to help', while in the other they received advice from an advisor who 'may sometimes try to mislead'. The order of advisors was counterbalanced between participants.

The advice offered by the advisors was drawn from equivalent distributions for both advisors, meaning that there were no systematic differences in the quality of the advice. Both advisors offered advice sampled from a roughly normal distribution around the correct answer with a standard deviation of 11 years.

Experiments

There were a series of experiments in this topic during which the experimental design was tweaked in order to find an effective manipulation and operationalisation of the key concepts. The experimental code for previous versions can be found on the commit history on GitHub for the main experimental file, as well as its dependencies.

To give a brief overview of the major changes:

- Version 2.0.0
  - introduced a much clearer manipulation
    - reminded participants of advisor description in a message they had to acknowledge
    - kept participant's group visible throughout
    - included a single trial where the misleading advisor did actually mislead the participant
- Version 2.1.0
  - removed actual differences in advice and adjusted advisor descriptions to match
- Version 3.0.0
  - added a question probing the perceived honesty of the advice between receiving advice and providing a final decision
- Version 3.0.1

- pre-registered replication of version 3.0.0
- Deviations from pre-registration:
  * Added a new exclusion rule for those people who use translation software.
  * Added exclusions for participants with NA values in the in- vs out-group t-test (where e.g. no outgroup advice was rated as 'honest').
  * Included frequentist stats in the trustworthiness questionnaire item t-test.
  * Fixed some labels on graphs
  * Exploratory analyses expanded to include analysis of trials with WoA > .05

Below, only the results from **version 3.0.1** are reported, but the results of this version are compatible with findings of previous versions. Data for all versions are accessible !TODO[archive data].

7.1.2 Results

!TODO[consistent advisor naming, including in figures]

Exclusions

Participants (total $n = 20$) could be excluded for a number of reasons: failing attention checks, having fewer than 11 trials which took less than 1 minute to complete (one participant), providing final decisions which were the same as the initial estimate on more than 90% of trials, or using non-English labels for the honesty questionnaire (one participant). The latter exclusion was added after data were collected because it was not anticipated that participants would use translation software in the task.

The final participant list consists of 18 participants who completed an average of 11.83 trials each.

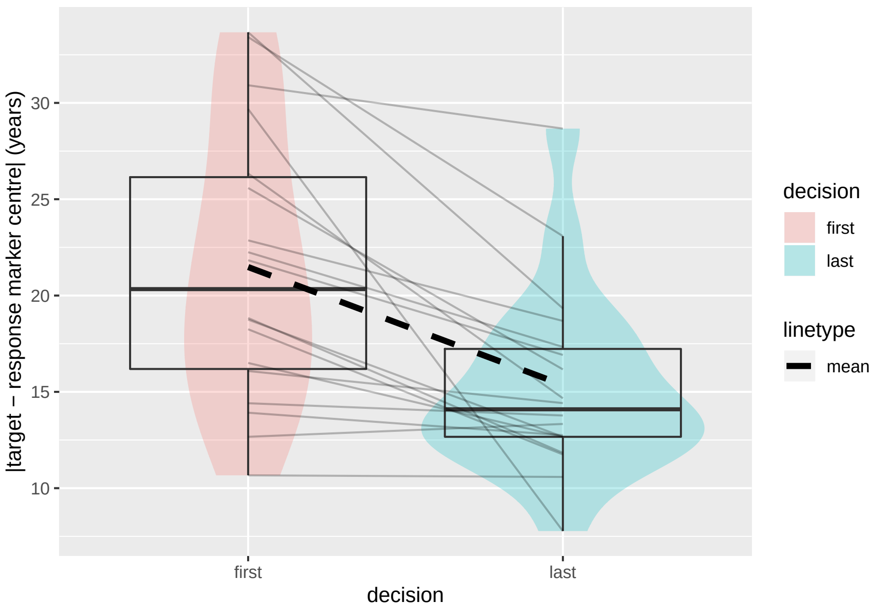

Figure 7.2: Mean answer error for initial estimates and final decisions. Faint lines show means for individual participants collapsed across trials, while boxplots, violins, and the heavy dashed line show aggregate participant means.

Task performance

Participants performed as expected, decreasing the error between the midpoint of their answer and the true answer from the initial estimate to the final decision, which suggests that they incorporated the advice, which was indicative of the correct answer.

The advice participants received was subject to variation, but there did not appear to be overall systematic differences between the advisors either in terms of error (distance between the centre of the advice estimate and the correct answer) or distance (distance between the centre of the advice estimate and the centre of the initial estimate).

!TODO[error and distance graphs]

Hypotheses

The hypotheses tested were that * weight on advice would be higher for advice rated as 'honest' versus advice rated as 'deceptive' * weight on advice would be higher for advisors who were described as 'never misleading', even for advice rated as 'honest'

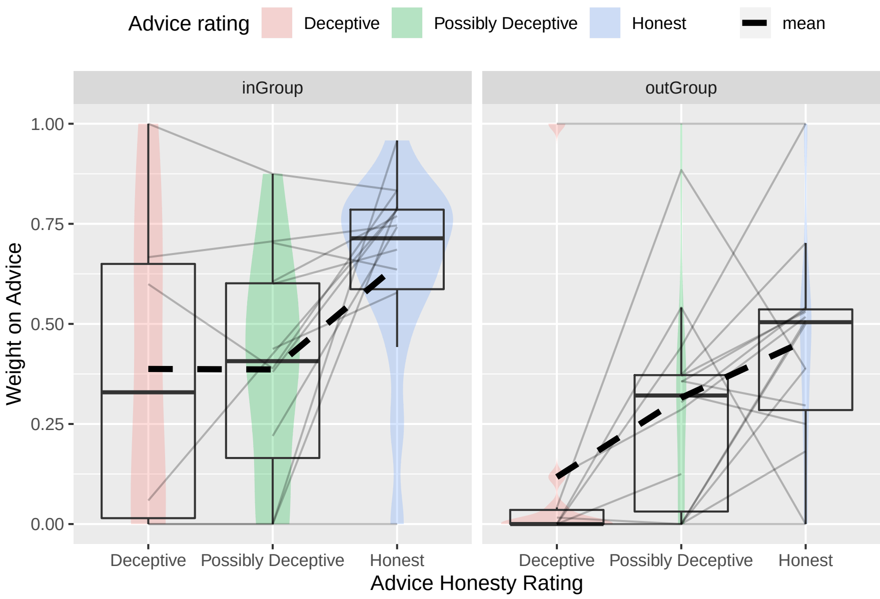

Figure 7.3: Influence of advice by advisor according to participants' perception of advice honesty. Faint lines show means for individual participants (missing segments indicate the participant rated no advice from that advisor with the designated category). Violins, box plots, and heavy dashed lines aggregate means for all participants.

Effect of advice As expected, participants were substantially more influenced by advice they rated as 'honest' compared to advice they rated as 'deceptive' ($t(10)$ = -5.01, p < .001, d = 1.55, BF = 69.66; M|deceptive = 0.18 [-0.04, 0.41], M|honest = 0.62 [0.47, 0.77]).

Effect of advisor Participants were also more influenced by 'honest' advice from the advisor who was described as 'never misleading' ($t(15)$ = 2.63, p = .019, d = 0.68, BF = 3.22; M|inGroup = 0.64 [0.51, 0.77], M|outGroup = 0.46 [0.30, 0.61]).

!TODO[plot of woa by advice rating and advisor]

Results plot The details of the data are best presented in a combined figure.

The figure shows that there is a different pattern in the responses to advice across advice ratings depending upon the advisor supplying the advice. Advice perceived as not being honest is discounted relatively equally from the advisor described as 'always helpful', and is still fairly influential in the final decision. The advice perceived as 'honest' from this advisor is the most influential of all advice. Likewise, the 'honest' advice from the advisor described as 'sometimes-misleading' is more influential than the 'possibly deceptive' advice, which is in turn more influential than the 'deceptive' advice. The 'deceptive' advice from the 'sometimes-misleading' advisor is almost completely ignored.

Within this broader pattern there is some variability between participants, including some participants whose responses violate the pattern quite dramatically - note the two participants with large decreases in influence for 'honest' advice from the 'sometimes-misleading' advisor as compared to the 'possibly deceptive' advice from the same. This may be a consequence of participants' values in each cell in the design being comprised of a low number of trials.

Exploration

!TODO[below] * were participants' criteria for judging advice as dishonest different between advisors?

```
* Type 1 SDT might be able to do this if a sensible definition of objectively-decep

    * E.g. whether averaging with the advice would reduce error

    * There will be all sorts of problems with this, noteably it'll classify 'never

* May be able to investigate this with probabilities of classification for several

    * what to do with 'possibly deceptive'?
```

7.1.3 Discussion

- Both the source and the plausibility of advice matter, and they seem to interact in a complex way.
- Flexible adaptation to advice context, both in terms of categorisation and influence.

- Support for the premise of the evolutionary simulations.
    - Does not prove people actually do have these hyperpriors

Limitations

- N
- No gender breakdown
    - simplicity/privacy vs data gap
- Difficult task (high levels of advice-taking by default)

7.2 Noise in the advice

The second scenario explored in the evolutionary models added noise to the advice agents received and demonstrated that this provided an evolutionary pressure towards egocentric discounting. This scenario is not explored in behavioural experiments because its conclusions are well supported by existing literature. Specifically, the literature demonstrates that the normativity of discounting where the judge outperforms the advisor, that advice-taking is sensitive to advisor expertise, and that people are likely to consider themselves superior to the average advisor.

The addition of noise in a point-value estimation task lowers the relative performance, and thus this scenario was essentially a manipulation of advisor expertise. Normative models support egocentric discounting where the judge is systematically more accurate than the advisor !TODO[PAR model paper, early advice-taking theory papers].

The decrease in advice-taking for novice as opposed to expert advisors is a robust result in the literature: !TODO[expertise manipulation experiment lit review].

The evidence above makes an empirical investigation on this point somewhat redundant, provided it can be shown that people have a naive assumption of superiority on a given task compared to the average other they are likely to encounter. This point is also well supported in the literature on self-serving biases !TODO[describe driving self-assessment, naive realism?, etc.]. It is expected that this is somewhat dependent upon the difficulty of the task presented; people faced with a difficult task will under- rather than overestimate their ability relative to others !TODO[hard-easy effect of advice].

Taken together, the above arguments demonstrate that, on average, people are likely to consider themselves more able on a given task than an arbitrary advisor, and consequently that they are likely to downweight advice relative to their own initial estimate. This behaviour is supported by normative mathematical models which show biasing towards the better estimator (in this case the judge) is the optimal strategy.

As discussed in the scenario, the belief that one is better at a task than the average advisor may not be misguided: advisors may not dedicate the same amount of time, concentration, or thought to producing advice as judges do for initial estimates. Judges have to live with the consequences of their decisions, while advisors do not.

7.3 Confidence mapping

The third scenario explored in the evolutionary models assigned each agent a confidence mapping, and demonstrated that discounting emerged as an appropriate response where the advisor's confidence mapping was unknown.

The key difficulty in conducting behavioural experiments to test the effects of known vs unknown confidence mapping is finding a manipulation of confidence mapping knowledge which is not confounded by familiarity with an advisor or the amount of information provided by an advisor.

We train people in two contexts, in both of which they receive advice with confidence from one of two advisors and see feedback on their final answers. In the first context, all the advice is well calibrated, and both advisors use either high or low confidence distributions. In the second context, one advisor uses a high and the other a low confidence distribution. Participants are then tested with a brief exposure to an advisor whose confidence distribution is drawn from the middle of the scale. Where participants are unprepared for different confidence mappings (the first context), they will treat the new advisor like the old advisors - considering the advice as high- or low-confidence (and weighting it accordingly) depending on whether it is high or low for the advisors' distributions. Where participants are prepared for different confidence mappings (the second context), they will treat the new advice as medium confidence, and weight it accordingly.

There will be a difference in the weight on advice in the probe trials between contexts: WoA will be highest where both trianing advisors have a low-confidence distribution, middling where they have different distributions, and lowest where they both have a high-confidence distribution.

Part IV

Interaction

8
Interaction of psychological processes across minds

Models say this will go horribly wrong.

9

Network effects of interaction

My models also suggest horrible things.

10
Real-world network effects

My models may not be good models.

Part V

Conclusion

11
Conclusion

I haven't wasted 3 years of my life and Nick's time.

Part VI

R Markdown demo

12

R Markdown Basics: The Markdown syntax

Contents

Here is a brief introduction to using *R Markdown*. *Markdown* is a simple formatting syntax for authoring HTML, PDF, and MS Word documents and much, much more. *R Markdown* provides the flexibility of *Markdown* with the implementation of **R** input and output. For more details on using *R Markdown* see `http://rmarkdown.rstudio.com`.

Be careful with your spacing in *Markdown* documents. While whitespace largely is ignored, it does at times give *Markdown* signals as to how to proceed. As a habit, try to keep everything left aligned whenever possible, especially as you type a new paragraph. In other words, there is no need to indent basic text in the Rmd document (in fact, it might cause your text to do funny things if you do).

12.1 Markdown basic syntax

12.1.1 Italics and bold

- *Italics* are done like *this* or _this_
- **Bold** is done like **this** or __this__
- ***Bold and italics*** is done like ***this***, ___this___, or (the most transparent solution) **_this_**

12.1.2 Inline code

- `Inline code` is created with backticks like `` `this` ``

12.1.3 Sub and superscript

Sub$_2$ and super^2 script is created like this~2~ and this^2^

12.1.4 Strikethrough

- ~~Strikethrough~~ is done ~~like this~~

12.1.5 'Escaping' (aka "What if I need an actual asterisk?")

- To include an actual *, _ or \, add another \ in front of them: *, _, \\

12.1.6 Endash (–), emdash (—)

- – and — with - -- and ---

12.1.7 Blockquotes

Do like this:

> Put a > in front of the line.

12.1.8 Headings

- are done with #'s of increasing number, i.e.
    - # First-level heading
    - ## Second-level heading
    - ### Etc.

In PDF output, a level-five heading will turn into a paragraph heading, i.e. `\paragraph{My level-five heading}`, which will appears as bold text on the same line as the subsequent paragraph.

12.1.9 Lists

Unordered list by starting a line with an * or a -:

- Item 1
- Item 2

Ordered lists by starting a line with a number:

1. Item 1
2. Item 2

Notice that you rcan mislabel the numbers and *Markdown* will still make the order right in the output.

To create a sublist, indent the values a bit (at least four spaces or a tab):

1. Item 1
2. Item 2
3. Item 3
    - Item 3a
    - Item 3b

12.1.10 Line breaks

The official *Markdown* way to create line breaks is by ending a line with more than two spaces.

Roses are red. Violets are blue.

This appears on the same line in the output, because we didn't add spaces after red.

Roses are red.
Violets are blue.

This appears with a line break because I added spaces after red.

I find this is confusing, so I recommend the alternative way: Ending a line with a backslash will also create a linebreak:

Roses are red.
Violets are blue.

To create a new paragraph, you put a blank line.

Therefore, this line starts its own paragraph.

12.1.11 Hyperlinks

- This is a hyperlink created by writing the text you want turned into a clickable link in `[square brackets followed by a](https://hyperlink-in-parentheses)`

12.1.12 Footnotes

- Are created[1] by writing either ^[my footnote text] for supplying the footnote content inline, or something like `[^a-random-footnote-label]` and supplying the text elsewhere in the format shown below [2]:

```
[^a-random-footnote-label]: This is a random test.
```

[1] my footnote text
[2] This is a random test.

12.1.13 Comments

To write comments within your text that won't actually be included in the output, you use the same syntax as for writing comments in HTML. That is, <!-- this will not be included in the output -->.

12.1.14 Math

The syntax for writing math is stolen from LaTeX. To write a math expression that will be shown **inline**, enclose it in dollar signs. - This: \$A = \pi*r^{2}\$ Becomes: $A = \pi * r^2$

To write a math expression that will be shown in a block, enclose it in two dollar signs.
This: \$\$A = \pi*r^{2}\$\$

Becomes:

$$A = \pi * r^2$$

To create numbered equations, put them in an 'equation' environment and give them a label with the syntax `(\#eq:label)`, like this:

```
\begin{equation}
  f\left(k\right) = \binom{n}{k} p^k\left(1-p\right)^{n-k}
  (\#eq:binom)
\end{equation}
```

Becomes:

$$f\left(k\right) = \binom{n}{k} p^k\left(1-p\right)^{n-k} \tag{12.1}$$

For more (e.g. how to theorems), see e.g. the documentation on bookdown.org

12.2 Additional resources

- *R Markdown: The Definitive Guide* - `https://bookdown.org/yihui/rmarkdown/`
- *R for Data Science* - `https://r4ds.had.co.nz`

13
Adding code

Contents

The magic of R Markdown is that we can add code within our document to make it dynamic.

We do this either as *code chunks* (generally used for loading libraries and data, performing calculations, and adding images, plots, and tables), or *inline code* (generally used for dynamically reporting results within our text).

13.1 Code chunks

The syntax of a code chunk is shown in Figure 13.1.

Common chunk options include (see e.g. bookdown.org):

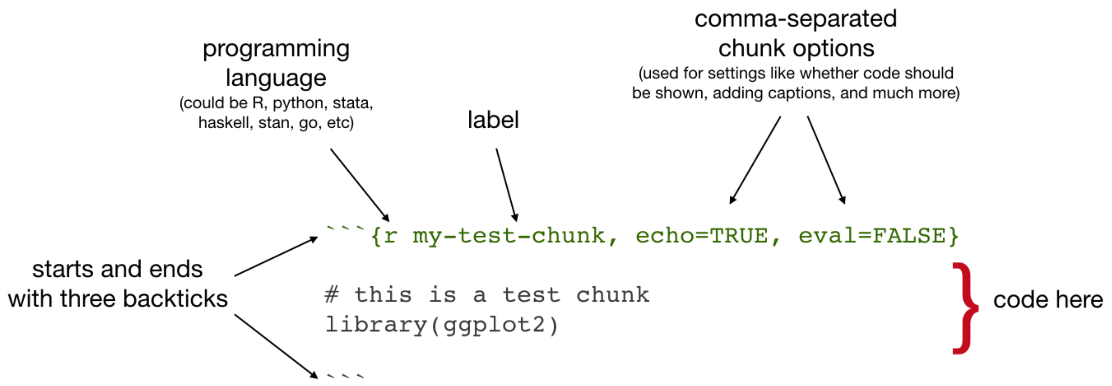

Figure 13.1: Code chunk syntax

- `echo`: whether or not to display code in knitted output
- `eval`: whether or to to run the code in the chunk when knitting
- `include`: wheter to include anything from the from a code chunk in the output document
- `fig.cap`: figure caption

IMPORTANT: Do *not* use underscoores in your chunk labels - if you do, you are likely to get an error in PDF output saying something like "! Package caption Error: \caption outside float".

13.1.1 Setup chunks

An R Markdown document usually begins with a chunk that is used to **load libraries**, and to **set default chunk options** with `knitr::opts_chunk$set`.

In your thesis, this will probably happen in **index.Rmd** and/or as opening chunks in each of your chapters.

```
‘‘‘{r setup, include=FALSE}
# don’t show code unless we explicitly set echo = TRUE
knitr::opts_chunk$set(echo = FALSE)


library(tidyverse)
‘‘‘
```

Figure 13.2: Oxford logo

13.1.2 Including images

Code chunks are also used for including images, with `include_graphics` from the `knitr` package, as in Figure 13.2

Useful chunk options for figures include: - `out.width` (use with a percentage) for setting the image size - if you've got an image that gets waaay to big in your output, it will be constrained to the page width by setting `out.width = "100%"`

Figure rotation

You can use the chunk option `out.extra` to rotate images.

The syntax is different for LaTeX and HTML, so for ease we might start by assigning the right string to a variable that depends on the format you're outputting to:

Then you can reference that variable as the value of `out.extra` to rotate images, as in Figure 13.3.

Figure 13.3: Oxford logo, rotated

13.1.3 Including plots

Similarly, code chunks are used for including dynamically generated plots. You use ordinary code in R or other languages - Figure 13.4 shows a plot of the `cars` dataset of stopping distances for cars at various speeds (this dataset is built in to **R**).

Under the hood, plots are included in your document in the same way as images - when you build the book or knit a chapter, the plot is automatically generated from your code, saved as an image, then included into the output document.

13.1.4 Including tables

Tables are usually included with the `kable` function from the `knitr` package.

Table 13.1 shows the first rows of that cars data - read in your own data, then use this approach to automatically generate tables.

- Gotcha: when using `kable`, captions are set inside the `kable` function
- The `kable` package is often used with the `kableExtra` package

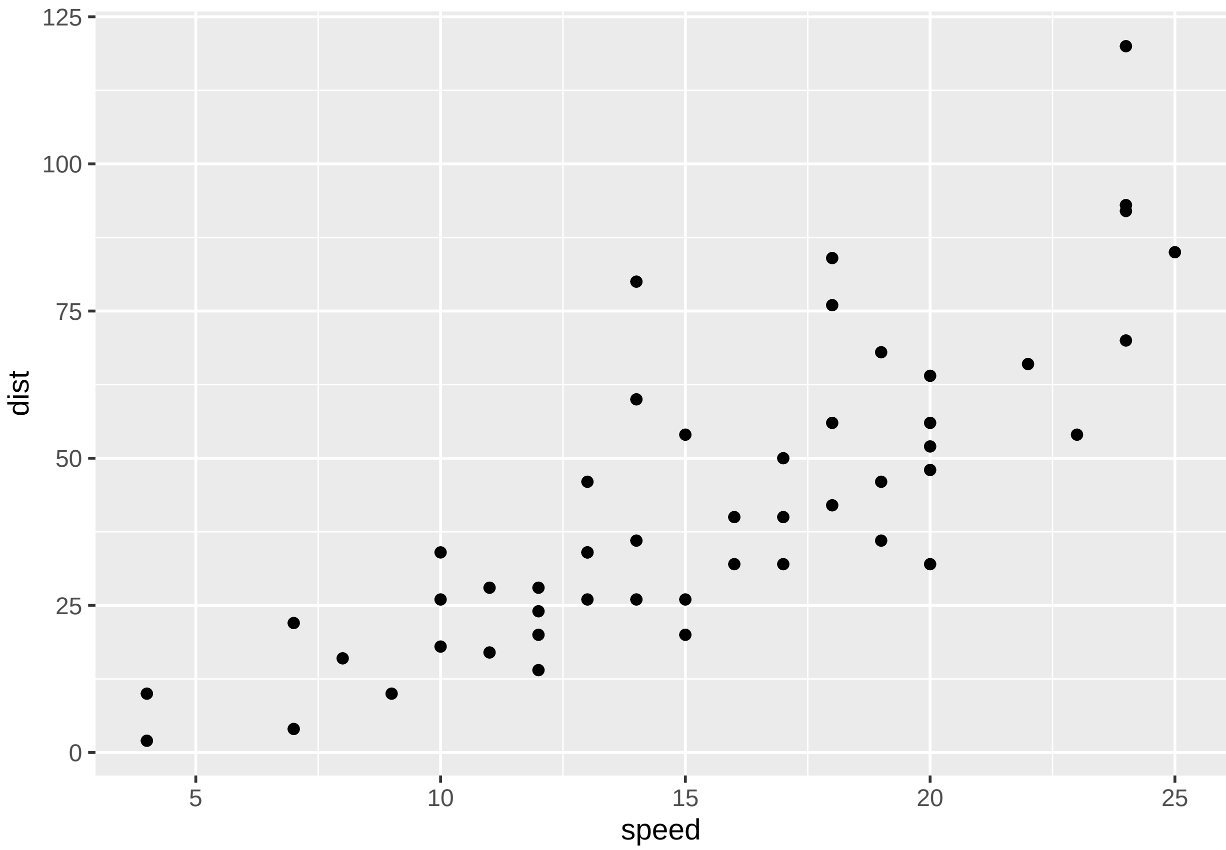

Figure 13.4: A ggplot of car stuff

Table 13.1: A knitr kable table

speed	dist
4	2
4	10
7	4
7	22
8	16
9	10

13.1.5 A note on content positioning

One thing that may be annoying is the way *R Markdown* handles "floats" like tables and figures.

In your PDF output, LaTeX will try to find the best place to put your object based on the text around it and until you're really, truly done writing you should just leave it where it lies.

When the time comes for you to make final tweaks to content positioning, read the relevant R Markdown documentation to see if there are easy ways to do what you want.

If you have very specific needs, you might have to read up on LaTeX (`https://en.wikibooks.org/wiki/LaTeX/Floats,_Figures_and_Captions`) for your PDF output and/or on how to style HTML documents with CSS for your gitbook output.

13.2 Inline code

'Inline code' simply means inclusion of code inside text.

The syntax for doing this is `` `r R_CODE` ``

For example, `` `r 4 + 4` `` would output 8 in your text.

You will usually use this in parts of your thesis where you report results - read in data or results in a code chunk, store things you want to report in a variable, then insert the value of that variable in your text.

For example, we might assign the number of rows in the `cars` dataset to a variable:

We might then write:
"In the `cars` dataset, we have `` `r num_car_observations` `` observations."

Which would output:
"In the `cars` dataset, we have 50 observations."

13.2.1 Referring to results computed in other languages than R

At the moment, inline code only works with R, so syntax such as `` `python code here` `` is not valid. However, you can use the `reticulate` package to access variables from python chunks. Here's a Python code chunk:

```
## 12
```

The `reticulate` package allows **R** to access variables defined in python environments with the syntax `py$variable`:

```
## Warning: package 'reticulate' was built under R version 3.5.3
## [1] 12
```

This means that inline, we can include results from python chunks with `` `r py$variable` ``. For example, we can state that the value of `my_number` defined in the python chunk, is 12.

14
Citations and cross-references

Contents

14.1 Citations

The usual way to include citations in an *R Markdown* document is to put references in a plain text file with the extension **.bib**, in **BibTex** format.[1] Then reference the path to this file in **index.Rmd**'s YAML header with `bibliography: example.bib`.

Most reference managers can create a .bib file with you references automatically. However, the **by far** best reference manager to use with *R Markdown* is Zotero with the Better BibTex plug-in, because the `citr` plugin for RStudio (see below) can read references directly from your Zotero library!

[1]The bibliography can be in other formats as well, including EndNote (**.enl**) and RIS (**.ris**), see rmarkdown.rstudio.com/authoring_bibliographies_and_citations.

Here is an example of an entry in a **.bib** file:

```
@article{Shea2014,
  author =          {Shea, Nicholas and Boldt, Annika},
  journal =         {Trends in Cognitive Sciences},
  pages =           {186--193},
  title =           {{Supra-personal cognitive control}},
  volume =          {18},
  year =            {2014},
  doi =             {10.1016/j.tics.2014.01.006},
}
```

In this entry highlighed section, 'Shea2014' is the **citation identifier**. To default way to cite an entry in your text is with this syntax: `[@citation-identifier]`.

So I might cite some things (**Shea2014**).

14.1.1 PDF output

In PDF output, the bibliography is handled by the OxThesis LaTeX template. If you set `bib-humanities: true` in **index.Rmd**, then in-text references will be formatted as author-year; otherwise references will be shown as numbers.

If you choose author-year formatting, a number of variations on the citation syntax are useful to know:

- Put author names outside the parenthesis
  - This: `@Shea2014 says blah.`
  - Becomes: **Shea2014** says blah.
- Include only the citation-year (in parenthesis)
  - This: `Shea et al. says blah [-@Shea2014]`
  - Becomes: Shea et al. says blah (**Shea2014**)
- Add text and page or chapter references to the citation
  - This: `[see @Shea2014, pp. 33-35; also @Wu2016, ch. 1]`

- Becomes: Blah blah (**Shea2014**; **Wu2016**).

14.1.2 Gitbook output

In gitbook output, citations are by default inserted in the Chicago author-date format.

To change the format, add `csl: some-other-style.csl` in **index.Rmd**'s YAML header. You can browse through and download styles at zotero.org/styles.

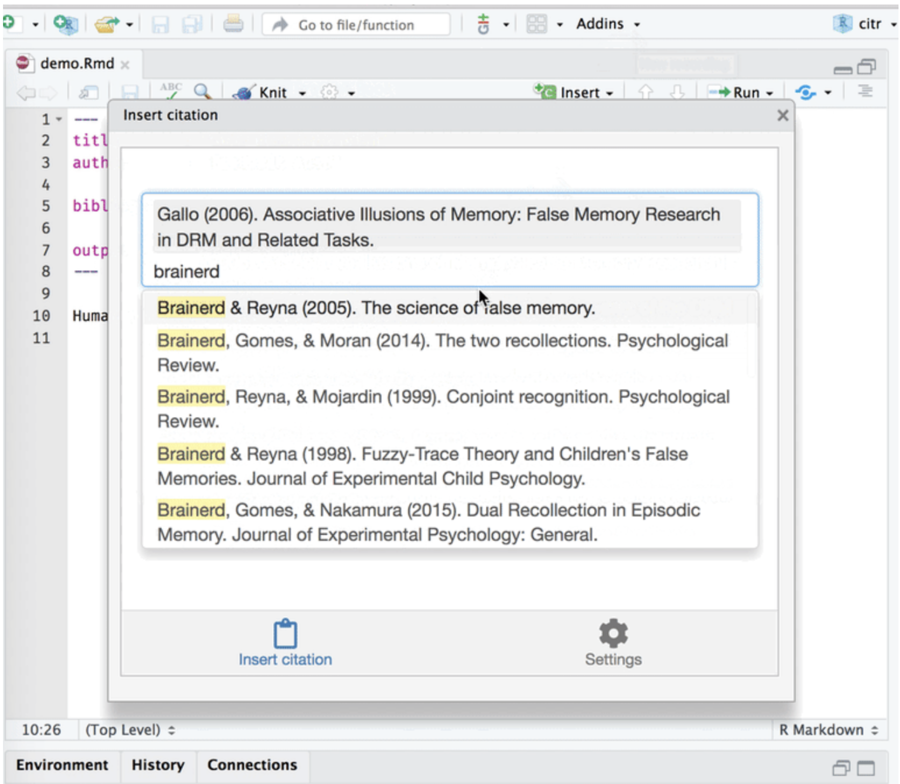

Figure 14.1: The 'citr' add-in

14.1.3 Insert references easily with the `citr` add-in

For an easy way to insert citations, try the `citr` RStudio add-in (Figure 14.1). You can install this add-in by typing `install.packages("citr")` in the R Console.

14.2 Cross-referencing

We can make cross-references to **sections** within our document, as well as to **figures** (images and plots) and **tables**.

The general cross-referencing syntax is **`\@ref(label)`**

14.2.1 Section references

Headers are automatically assigned a reference label, which is the text in lower caps separated by dashes. For example, `# My header` is automatically given the label `my-header`. So `# My header` can be referenced with `\@ref(my-section)`

Remember what we wrote in section 14.1?

We can also use **hyperlink syntax** and add # before the label, though this is only guaranteed to work properly in HTML output:

- So if we write `Remember what we wrote up in [the previous section](#citations)?`
- It becomes Remember what we wrote up in the previous section?

Creating custom labels

It is a very good idea to create **custom labels** for our sections. This is because the automatically assigned labels will change when we change the titles of the sections - to avoid this, we can create the labels ourselves and leave them untouched if we change the section titles.

We create custom labels by adding `{#label}` after a header, e.g. `# My section {#my-label}`. See our chapter title for an example. That was section 14.

14.2.2 Figure (image and plot) references

- To refer to figures (i.e. images and plots) use the syntax `\@ref(fig:label)`
- **GOTCHA**: Figures and tables must have captions if you wish to cross-reference them.

Let's add an image:

We refer to this image with `\@ref(fig:captain)`. So Figure 14.2 is this image.

And in Figure 13.4 we saw a cars plot.

14.2.3 Table references

- To refer to tables use the syntax `\@ref(tab:label)`

Let's include a table:

We refer to this table with `\@ref(tab:cars-table2)`. So Table 14.1 is this table.

And in Table 13.1 we saw more or less the same cars table.

Figure 14.2: A marvel-lous meme

Table 14.1: Stopping cars

speed	dist
4	2
4	10
7	4
7	22
8	16

14.2.4 Including page numbers

Finally, in the PDF output we might also want to include the page number of a reference, so that it's easy to find in physical printed output. LaTeX has a command for this, which looks like this: `\pageref{fig/tab:label}` (note: curly braces, not parentheses)

When we output to PDF, we can use raw LaTeX directly in our .Rmd files. So if we wanted to include the page of the cars plot we could write:

- This: `Figure \@ref(fig:cars-plot) on page \pageref(fig:cars-plot)`
- Becomes: Figure 13.4 on page 65

Include page numbers only in PDF output

A problem here is that LaTeX commands don't display in HTML output, so in the gitbook output we'd see simply "Figure 13.4 on page".

One way to get around this is to use inline R code to insert the text, and use an `ifelse` statement to check the output format and then insert the appropriate text.

- So this: `` `r ifelse(knitr::is_latex_output(), "Figure \\@ref(fig:cars-plot) on page \\pageref{fig:cars-plot}", "")` ``
- Inserts this (check this on both PDF and gitbook): Figure 13.4 on page 65

Note that we need to escape the backslash with another backslash here to get the correct output.

15 Final Notes on The OxThesis template and on collaboration

15.1 Beginning chapters with quotes

The OxThesis LaTeX template lets you inject some wittiness into your thesis by including a block of type `savequote` at the beginning of chapters. To do this, use the syntax ` ```{block type='savequote'} `.[1]

Add the reference for the quote with the chunk option `quote_author="my author name"`. You will also want to add the chunk option `include=knitr::is_latex_output()` so that quotes are only included in PDF output.

It's not possible to use markdown syntax inside chunk options, so if you want to e.g. italicise a book name in the reference use a 'text reference': Create a named piece of text with '(ref:label-name) My text', then point to this in the chunk option with `quote_author='(ref:label-name)'`.

15.2 Highlighting corrections

For when it comes time to do corrections, you may want to highlight changes made when you submit a post-viva, corrected copy to your examiners so they can quickly

[1] For more on custom block types, see the relevant section in *Authoring Books with R Markdown*.

verify you've completed the task. You can do so like this:

15.2.1 Short, inline corrections

Highlight **short, inline corrections** by wrapping them in a `<span>` tag with the class 'correction'. In other words, if you do `<span class="correction">like this</span>`, the text between the span tags will be highlighted in blue in the output.

15.2.2 Blocks of added or changed material

Highlight entire **blocks of added or changed material** by putting them in a block of type `correction`, using the syntax ` ```{block type='correction'}`.[2] Like so:

15.2.3 Stopping corrections from being highlighted in the output

For **PDF** output, go to **index.Rmd** and (i) set `corrections: false` under `params` in the YAML header (stops block of corrections from being highlighted), (ii) comment out `pandoc_args: ["--lua-filter=scripts_and_filters/correction_filter.lua"]` (stops inline corrections from being highlighted).

For **gitbook** output, go to **style.css** and comment out the styling for `.correction`.

15.3 Diving in to the OxThesis LaTeX template

For LaTeX minded people, you can read through **templates/template.tex** to see which additional customisation options are available as well as **templates/ociamthesis.cls** which supplies the base class. For example, **template.tex** provides an option for master's degree submissions, which changes identifying information to candidate number and includes a word count. At the time of writing, you must set this directly in **template.tex** rather than from the YAML header in **index.Rmd**.

[2] In the **.tex** file for PDF output, this will put the content between `\begin{correction}` and `\end{correction}`; in gitbook output it will be put between `<div class="correction">` and `</div>`.

15.4 Collaborative writing

Best practices for collaboration and change tracking when using R Markdown are still an open question. In the blog post **One year to dissertate** by Lucy D'Agostino, which I highly recommend, the author notes that she knits .Rmd files to a `word_document`, then uses the `googledrive` R package to send this to Google Drive for comments / revisions from co-authors, then incorporates Google Drive suggestions *by hand* into the .Rmd source files. This is a bit clunky, and there are ongoing discussions among the *R Markdown* developers about what the best way is to handle collaborative writing (see issue #1463 on GitHub, where CriticMarkup is among the suggestions).

For now, this is an open question in the community of R Markdown users. I often knit to a format that can easily be imported to Google Docs for comments, then go over suggested revisions and manually incorporate them back in to the .Rmd source files. For articles, I sometimes upload a near-final draft to Overleaf, then collaboratively make final edits to the LaTeX file there. I suspect some great solution will be developed in the not-to-distant future, probably by the RStudio team.

Conclusion

If we don't want Conclusion to have a chapter number next to it, we can add the `{-}` attribute.

More info

And here's some other random info: the first paragraph after a chapter title or section head *shouldn't be* indented, because indents are to tell the reader that you're starting a new paragraph. Since that's obvious after a chapter or section title, proper typesetting doesn't add an indent there.

Appendices

A
The First Appendix

This first appendix includes an R chunk that was hidden in the document (using `echo = FALSE`) to help with readibility:

In 02-rmd-basics-code.Rmd

```
library(tidyverse)
knitr::include_graphics("figures/chunk-parts.png")
```

And here's another one from the same chapter, i.e. Chapter 13:

```
knitr::include_graphics("figures/beltcrest.png")
```

B
The Second Appendix, for Fun

Works Cited

Ais, Joaquín et al. (Jan. 2016). “Individual Consistency in the Accuracy and Distribution of Confidence Judgments”. In: *Cognition* 146.Supplement C, pp. 377–386.

Azaria, Amos et al. (Jan. 2016). “Strategic Advice Provision in Repeated Human-Agent Interactions”. English. In: *Autonomous Agents and Multi-Agent Systems* 30.1. WOS:000370153000002, pp. 4–29.

Bahrami, Bahador et al. (2010). “Optimally Interacting Minds”. In: *Science* 329.5995, pp. 1081–1085.

Barberá, Pablo (2015). “How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S.” en. In: p. 46.

Byrne, Kaileigh A. et al. (Oct. 2016). “Ostracism Reduces Reliance on Poor Advice from Others during Decision Making”. English. In: *Journal of Behavioral Decision Making* 29.4. WOS:000383623700006, pp. 409–418.

Cardoso, Felipe Maciel et al. (July 2017). “Topical Homophily in Online Social Systems”. en. In:

Colleoni, Elanor, Alessandro Rozza, and Adam Arvidsson (Apr. 2014). “Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data: Political Homophily on Twitter”. en. In: *Journal of Communication* 64.2, pp. 317–332.

Dana, Jason and Daylian M Cain (Dec. 2015). “Advice versus Choice”. en. In: *Current Opinion in Psychology* 6, pp. 173–176.

Dandekar, Pranav, Ashish Goel, and David T. Lee (Apr. 2013). “Biased Assimilation, Homophily, and the Dynamics of Polarization”. en. In: *Proceedings of the National Academy of Sciences* 110.15, pp. 5791–5796.

Galton, Francis (1907). “Vox Populi (the Wisdom of Crowds)”. In: *Nature* 75, pp. 450–451.

Garrett, R. Kelly (Jan. 2009a). “Echo Chambers Online?: Politically Motivated Selective Exposure among Internet News Users”. en. In: *Journal of Computer-Mediated Communication* 14.2, pp. 265–285.

— (Dec. 2009b). “Politically Motivated Reinforcement Seeking: Reframing the Selective Exposure Debate”. en. In: *Journal of Communication* 59.4, pp. 676–699.

Gino, Francesca (Nov. 2008). “Do We Listen to Advice Just Because We Paid for It? The Impact of Advice Cost on Its Use”. en. In: *Organizational Behavior and Human Decision Processes* 107.2, pp. 234–245.

Gino, Francesca, Alison Wood Brooks, and Maurice E Schweitzer (2012). “Anxiety, Advice, and the Ability to Discem: Feeling Anxious Motivates Individuals to Seek and Use Advice”. en. In: *Journal of Personality and Social Psychology* 102.3, pp. 497–512.

Gino, Francesca and Don A. Moore (Jan. 2007). “Effects of Task Difficulty on Use of Advice”. en. In: *Journal of Behavioral Decision Making* 20.1, pp. 21–35.

Hertz, Uri and Bahador Bahrami (Apr. 2018). “Intrinsic Value of Social Influence over Others”. In:

Hertz, Uri, Stefano Palminteri, et al. (Dec. 2017). “Neural Computations Underpinning the Strategic Management of Influence in Advice Giving”. en. In: *Nature Communications* 8.1, p. 2191.

Hütter, Mandy and Fabian Ache (2016). “Seeking Advice: A Sampling Approach to Advice Taking”. en. In: *Judgment and Decision Making* 11.4, p. 16.

Kleiner, M, D Brainard, and D Pelli (2007). “What’s New in Psychtoolbox-3? Perception 36 ECVP Abstract Supplement”. In:

Kobayashi, Tetsuro and Ken’ichi Ikeda (Sept. 2009). “SELECTIVE EXPOSURE IN POLITICAL WEB BROWSING: Empirical Verification of ‘Cyber-Balkanization’ in Japan and the USA”. en. In: *Information, Communication & Society* 12.6, pp. 929–953.

Liberman, Varda et al. (Mar. 2012). “Naïve Realism and Capturing the “Wisdom of Dyads””. en. In: *Journal of Experimental Social Psychology* 48.2, pp. 507–512.

MATLAB (2017). The Mathworks, Inc.

Mayer, Roger C., James H. Davis, and F. David Schoorman (1995). “An Integrative Model of Organizational Trust”. In: *Academy of management review* 20.3, pp. 709–734.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook (2001). “Birds of a Feather: Homophily in Social Networks”. In: *Annual Review of Sociology* 27.1, pp. 415–444.

Minson, Julia A. and Jennifer S. Mueller (Mar. 2012). “The Cost of Collaboration: Why Joint Decision Making Exacerbates Rejection of Outside Information”. en. In: *Psychological Science* 23.3, pp. 219–224.

Navajas, Joaquin et al. (Nov. 2017). “The Idiosyncratic Nature of Confidence”. En. In: *Nature Human Behaviour* 1.11, p. 810.

Nelson, Jacob L. and James G. Webster (July 2017). “The Myth of Partisan Selective Exposure: A Portrait of the Online Political News Audience”. en. In: *Social Media + Society* 3.3, p. 205630511772931.

Pescetelli, Niccolo (2017). “On the Use of Metacognitive Signals to Navigate the Social World”. DPhil. University of Oxford.

Pescetelli, Niccolo and Nick Yeung (Sept. 2018). “The Role of Decision Confidence in Advice-Taking and Trust Formation”. In: *arXiv:1809.10453 [cs]*. arXiv: `1809.10453 [cs]`.

R Core Team (2018). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rader, Christina A., Richard P. Larrick, and Jack B. Soll (Aug. 2017). “Advice as a Form of Social Influence: Informational Motives and the Consequences for Accuracy”. en. In: *Social and Personality Psychology Compass* 11.8, n/a–n/a.

Richards, Andrew (Aug. 2015). “University of Oxford Advanced Research Computing”. In:

Ronayne, David and Daniel Sgroi (2018). “Ignoring Good Advice”. In:

Sah, Sunita, Don A. Moore, and Robert J. MacCoun (July 2013). “Cheap Talk and Credibility: The Consequences of Confidence and Accuracy on Advisor Credibility and Persuasiveness”. In: *Organizational Behavior and Human Decision Processes* 121.2, pp. 246–255.

Schul, Yaacov and Noam Peni (Oct. 2015). “Influences of Distrust (and Trust) on Decision Making”. English. In: *Social Cognition* 33.5. WOS:000364258500004, pp. 414–435.

Schultze, Thomas, Andreas Mojzisch, and Stefan Schulz-Hardt (May 2017). “On the Inability to Ignore Useless Advice: A Case for Anchoring in the Judge-Advisor-System”. en. In: *Experimental Psychology* 64.3, pp. 170–183.

Sears, David O. and Jonathan L. Freedman (1967). “Selective Exposure to Information: A Critical Review”. In: *Public Opinion Quarterly* 31.2, pp. 194–213.

See, Kelly E. et al. (Nov. 2011). “The Detrimental Effects of Power on Confidence, Advice Taking, and Accuracy”. In: *Organizational Behavior and Human Decision Processes* 116.2, pp. 272–285.

Sniezek, Janet A., Gunnar E. Schrah, and Reeshad S. Dalal (July 2004). “Improving Judgement with Prepaid Expert Advice”. en. In: *Journal of Behavioral Decision Making* 17.3, pp. 173–190.

Sniezek, Janet A. and Lyn M. Van Swol (Mar. 2001). “Trust, Confidence, and Expertise in a Judge-Advisor System”. In: *Organizational Behavior and Human Decision Processes* 84.2, pp. 288–307.

Soll, Jack B. and Albert E. Mannes (Jan. 2011). “Judgmental Aggregation Strategies Depend on Whether the Self Is Involved”. en. In: *International Journal of Forecasting* 27.1, pp. 81–102.

Song, Chen et al. (Dec. 2011). “Relating Inter-Individual Differences in Metacognitive Performance on Different Perceptual Tasks”. In: *Consciousness and Cognition*. From Dreams to Psychosis: A European Science Foundation Exploratory Workshop 20.4, pp. 1787–1792.

Song, Hyunjin and Hajo G. Boomgaarden (Apr. 2017). “Dynamic Spirals Put to Test: An Agent-Based Model of Reinforcing Spirals Between Selective Exposure, Interpersonal Networks, and Attitude Polarization”. en. In: *Journal of Communication* 67.2, pp. 256–281.

Sunstein, Cass R. (2002). *Republic.Com*. en. Princeton University Press.

— (Apr. 2018). *#Republic: Divided Democracy in the Age of Social Media*. en. Princeton University Press.

Trouche, Emmanuel et al. (Jan. 2018). “Vigilant Conservatism in Evaluating Communicated Information”. en. In: *PLOS ONE* 13.1. Ed. by Alexander N. Sokolov, e0188825.

Van Swol, Lyn M. (Jan. 2011). “Forecasting Another’s Enjoyment versus Giving the Right Answer: Trust, Shared Values, Task Effects, and Confidence in Improving the Acceptance of Advice”. en. In: *International Journal of Forecasting* 27.1, pp. 103–120.

Yaniv, I. (Jan. 2004). “Receiving Other People’s Advice: Influence and Benefit”. English. In: *Organizational Behavior and Human Decision Processes* 93.1. WOS:000188504000001, pp. 1–13.

Yaniv, Ilan and Shoham Choshen-Hillel (Dec. 2012). “Exploiting the Wisdom of Others to Make Better Decisions: Suspending Judgment Reduces Egocentrism and Increases Accuracy”. en. In: *Journal of Behavioral Decision Making* 25.5, pp. 427–434.

Yaniv, Ilan and Eli Kleinberger (Nov. 2000). “Advice Taking in Decision Making: Egocentric Discounting and Reputation Formation”. en. In: *Organizational Behavior and Human Decision Processes* 83.2, pp. 260–281.

Yaniv, Ilan and Maxim Milyavsky (May 2007). “Using Advice from Multiple Sources to Revise and Improve Judgments”. English. In: *Organizational Behavior and Human Decision Processes* 103.1. WOS:000246807900007, pp. 104–120.