



Thesis Title: Framing Effects and Context in Language Comprehension

Degree Title: Doctor of Philosophy

Department: Philosophy

Candidate Name: Sarah Fisher

Submission Date: April 2020

DECLARATION

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Sarah Fisher

ACKNOWLEDGEMENTS

The research was generously supported by the Arts and Humanities Research Council, via a studentship awarded to me by the South, West and Wales Doctoral Training Partnership (grant number AH\503939\1). I would like to express my sincere gratitude to Emma Borg and Nat Hansen for their invaluable advice and unflagging encouragement throughout this process; I genuinely could not have wished for a better supervisory team. Thank you also to Kathryn Francis for her guidance on running the experiments, to Paul Egré, Dan Harris, Marat Shardimgaliev, and Antonio Scarafone for extended discussions of the material in later chapters, to many other colleagues whose comments have helped shape and refine my ideas along the way, and – of course – to my long-suffering friends and family for their immense support.

ABSTRACT

Linguistic framing effects occur when audiences respond differently to the same information, just because of the wording the speaker used. For example, empirical research suggests that if a risky surgery is described in terms of the chance of *surviving*, people are more likely to go ahead than if it is described in terms of the chance of *dying* (Wilson, Kaplan, & Schneiderman, 1987). Advertisers, politicians, and rhetoricians have always intuitively understood that our judgements and decisions can be shaped by the words we use. Over the last fifty years, though, framing has become the subject of extensive scientific investigation, beginning with the publication in 1981 of 'The Framing of Decisions and the Psychology of Choice' by Amos Tversky and Daniel Kahneman. In the psychological literature, framing effects are standardly taken as evidence of humans' irrationality. My research challenges that interpretation. I identify a handful of semantic and pragmatic features of speakers' frames, which could explain why audiences form distinct representations of the situations being described. In other words, audiences may be tracking genuine differences in the meanings of alternative frames, rather than committing some sort of reasoning error. I also use framing research as an empirical case study, to inform philosophical understandings of the distinction between semantics and pragmatics. Turning to look beyond academic debates, I end by exploring speakers' and hearers' responsibilities for problematic framing in public discourse.

CONTENTS

| | |
|---|------------|
| Chapter 1. Framing: An Introduction | 5 |
| Chapter 2. Semantic Conceptions | 18 |
| Chapter 3. Inexhaustive Alternatives | 42 |
| Chapter 4. Lower-Boundedness | 66 |
| Chapter 5. Truth-Evaluability | 87 |
| Chapter 6. Information Leakage | 107 |
| Chapter 7. Implicating Reference Point Information | 129 |
| Chapter 8. Keeping Semantics Minimal | 153 |
| Chapter 9. Framing: The Fault Lines | 174 |
| Appendix | 198 |
| Bibliography | 199 |

CHAPTER 1

Framing: An Introduction

Abstract

In this chapter, I define the concepts of ‘framing’ and ‘framing effect’ which I will be working with throughout the thesis. I provide a brief historical overview of the psychological investigations into these phenomena, and I situate my own research within the literature. I close by outlining the arguments I will develop in subsequent chapters.

1. Introduction

The publication in 1981 of ‘The Framing of Decisions and the Psychology of Choice’, by Amos Tversky and Daniel Kahneman, marked the advent of psychological research into framing and framing effects. Since then a large literature has grown up around these topics, and it continues to expand and evolve. In its broadest sense, framing concerns the context in which information is presented. Of particular interest are pairs of communicative stimuli whose informational contents seem to be the same in all relevant respects; nevertheless, various contextual factors can systematically affect how the audience responds to them. In this thesis, I will restrict my attention to *linguistic* framing. I will focus on variations in the words in which information is packaged and conveyed. I will be concerned with pairs of uttered sentences whose informational contents seem to be logically equivalent to one another.¹ I call such pairs of sentences ‘alternative frames’. The framing effects I am interested in are systematic differences in audiences’ responses, which are brought about by the use of one or other of a pair of alternative frames in an otherwise identical context. Reflecting the scope and coverage of the experimental literature, the audience responses I will focus on are elicited judgements and decisions.²

The structure of this introductory chapter is as follows: in §2 I outline the normative principles of the classical ‘expected utility theory’ of decision-making under risk. In §3 I describe the ‘prospect theory’ developed by Kahneman and Tversky, which is put forward as a descriptively adequate alternative. In §4 I show how prospect theory is extended to account for risky-choice framing effects. In §5 I focus on the theory’s assumption that recipients of alternative frames represent prospects differently. In §6 I explain why their doing so is standardly considered to be

¹ Using the terminology introduced by Druckman (2004, 2011), I will therefore be concerned with ‘equivalency framing’ rather than ‘emphasis framing’. The latter appeals to substantively distinct dimensions of evaluation – for example, one might frame climate change as an *environmental* issue or a *social justice* issue. Frisch (1993) makes a similar distinction between framing effects considered ‘strictly’ or ‘loosely’.

² As will become clearer below, I will primarily be interested in the different mental representations hearers form in response to alternative frames, which mediate these behavioural responses.

irrational. In §7 I home in on my key research question, which concerns how – and why – alternative frames induce distinct representations. In §8 I introduce the paradigm of ‘attribute framing’. In §9 I provide an overview of the thesis.

2. Expected Utility Theory

Psychological research on framing began as part of a wider challenge to the classical rational-choice model of decision-making known as ‘expected utility theory’. Expected utility theory (EUT) comprises of a set of axioms specifying how rational agents make decisions under conditions of risk. Decision-making under risk involves evaluating ‘prospects’, which are sets of possible outcomes. Each possible outcome is, in turn, associated with a level of utility and a probability of occurring, ranging from zero (impossible) to one (certain).³ To give a concrete example, one might (if sufficiently lucky) be faced with a choice between (i) receiving \$200 for sure, or (ii) having a 50% chance of receiving \$400 and a 50% chance of receiving \$0. This scenario involves two prospects: the first prospect comprises a single outcome of receiving \$200 with a probability of 1. The second prospect comprises two possible outcomes – receiving \$400 or receiving \$0 – each of which is associated with a probability of 0.5.

According to EUT, the ‘expected utility’ of any prospect is a weighted sum of the expected utilities of each possible outcome. The expected utility of each *outcome* is the product of its utility and its probability of occurring. So, the expected utility of our first prospect above would be $\$200 \times 1 = \200 .⁴ The expected utility of the second prospect is also \$200. We arrive at this in a different way; namely, by multiplying each of the two possible outcomes (winning \$400 and winning \$0) by the probabilities of their occurring (0.5 for both outcomes) and then summing the results: $(\$400 \times 0.5 = \$200) + (\$0 \times 0.5 = \$0) = \$200$.

Most versions of EUT assume that a prospect in which the outcome is sure to obtain should be preferred to a ‘risky’ prospect of equal expected utility. Therefore, even though both of our prospects have an expected utility of \$200, EUT standardly assumes that receiving \$200 for sure should be preferred to the ‘risky’ prospect of receiving \$400 or \$0. In other words, EUT standardly assumes that rational agents are *risk-averse*.

³ In contrast with decision-making under *risk*, decision-making under *uncertainty* involves evaluating underspecified possibilities. For example, these possibilities may not be jointly exhaustive, and the probability of their occurring may not be known. The distinction between risk and uncertainty is important to those who argue that each demands a radically different decision-making strategy – see, for example (Gigerenzer & Gaissmaier, 2011). For current purposes I leave uncertainty to one side and limit my focus to decision-making under *risk*.

⁴ For simplicity, I will equate utility with monetary value, so that the utility associated with receiving \$200 just is \$200.

3. Prospect Theory

Kahneman and Tversky (1979) bring together a large body of experimental evidence which suggests that human decision-makers systematically violate key tenets of EUT. Of most relevance for our purposes is their finding that people tend to reason about a prospect differently depending on whether its expected utility is positive or negative – in other words, depending on whether it represents a gain or loss in utility. Specifically, while people tend to prefer sure gains to risky gains (as is standardly assumed by EUT), they tend to prefer risky losses to sure losses. For example, if offered a choice between (i) *losing* \$200 for sure, and (ii) a 50% chance of *losing* \$400 and a 50% chance of *losing* \$0, people tend to prefer option (ii). Thus, the empirical evidence suggests that human agents are *risk-averse* in the domain of gains but *risk-seeking* in the domain of losses.⁵ Kahneman and Tversky dub the sensitivity of people's risk attitudes to gains or losses the 'reflection effect'. The label reflects the idea that attitudes to risk are reversed by reflecting prospects around the zero utility point.

In order to accurately describe human decision-making under risk, Kahneman and Tversky develop a model of decision-making under risk known as 'prospect theory'. On this model, the value assigned to a prospect will depend in part on whether it constitutes a gain or a loss in utility. This allows the theory to predict risk-averse behaviour in the domain of gains and risk-seeking behaviour in the domain of losses.⁶

4. From Reflection to Framing

Crucially, prospect theory is extended to cases where the prospects merely *seem* like gains or losses, despite their having the same overall impact on utility. These cases arise where decision-makers have different reference points in mind when considering the prospects. As Kahneman and Tversky put it:

The reference point usually corresponds to the current asset position, in which case gains and losses coincide with the actual amounts that are received or paid. However, the location of the reference point, and the consequent coding of outcomes as gains or

⁵ The picture is further complicated by evidence that the subjective value of an outcome is not a straightforward product of its utility and its probability; instead, there are differences in how utilities and probabilities are integrated into the decision procedure depending on their respective magnitudes. Moreover, these differences are affected by whether the prospect is a gain or a loss. For example, people prefer highly improbable large gains to sure small gains; but they prefer sure small losses to highly improbable large losses. These details are largely tangential to my concerns so I will set them aside for now.

⁶ Prospect theory also includes weighting components that capture sensitivities to the magnitudes of the utilities and probabilities of possible outcomes (as discussed in the previous footnote). Again, I will ignore these aspects of the model here. However, see (Kühberger, Schulte-Mecklenbeck, & Perner, 1999) for further analysis of how far they are supported by data from 'risky choice' framing experiments, like the 'Asian Disease Problem' discussed next.

losses, can be affected by the formulation of the offered prospects, and by the expectations of the decision maker. (Kahneman & Tversky, 1979, p. 274)

One way of making prospects seem like gains or losses is by describing them in different words – or using *alternative frames*. A prime example is provided by the ‘Asian Disease Problem’ (ADP). Originally used in an experiment reported by Tversky and Kahneman (1981), variations of this problem permeate the subsequent framing literature so thoroughly that it is now something of a classic of the genre. In the original experiment, Tversky and Kahneman presented participants with the following hypothetical scenario:

Imagine the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume the exact scientific estimates of the consequences are as follows:

In one condition participants received the following options:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.

In another condition they received the following:

If Program C is adopted, 400 people will die.

If Program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.

Participants in each condition were then asked which of the options they would favour.

Tversky and Kahneman found that 72% of participants in the first, ‘be saved’, condition favoured Program A, compared with 28% who favoured Program B. Meanwhile, in the second, ‘die’, condition, 78% of participants favoured Program D, compared with 22% who favoured Program C. The pattern of responses is considered to be puzzling because Programs A and C are standardly taken to describe exactly the same sure prospect (of 200 people being saved and 400 dying). Similarly, Programs B and D are standardly taken to describe the same risky prospect (of a one-third probability of 600 people being saved with none dying, and a two-thirds probability of 600 dying with none being saved).⁷ The only difference between the two framing conditions, then, is whether the options are worded positively, with the outcomes expressed in terms of the numbers of people who will ‘be saved’, or negatively, in terms of the numbers of people who will

⁷ In later chapters I will challenge these apparent equivalences in various ways. However, let’s assume for now that the framing conditions straightforwardly describe the same pair of options.

'die'.⁸ Therefore this is not a reflection effect but a framing effect: whereas reflection effects concern *actual* gains and losses (as in the earlier examples of gaining money and losing money), framing effects concern merely *apparent* gains and losses (as in the ADP, where the actual numbers of people living and dying remain exactly the same).⁹

Even assuming that risk attitudes vary across gain and loss domains, the responses to the ADP remain puzzling: both pairs of options represent the same change in utility, so why didn't participants consistently prefer either the sure or the risky prospect? The results demonstrate a reversal in preferences across framing conditions, with most participants in the 'be saved' condition favouring the sure option, and most participants in the 'die' condition favouring the risky option. It seems, then, that a mere difference in wording – or framing – affects the values that decision-makers assign to the prospects.

Similar risky-choice framing effects have been observed in myriad subsequent studies, using a variety of scenarios and experimental designs (for surveys of the literature, see (Kühberger, 1998; Kühberger et al., 1999; Levin, Schneider, & Gaeth, 1998; Steiger & Kühberger, 2018)). Note that, although not all studies find wholesale *reversals* in preferences, there is a very robust statistically significant *shift* in the expected direction, i.e. towards the sure option in the positively worded framing condition and towards the risky option in the negatively worded framing condition (Kühberger, 1998, p. 30; Levin et al., 1998, p. 153). I will understand framing effects to encompass shifts as well as wholesale reversals. This is in line with definitions adopted by many others in the literature, including Levin et al. (1998) and Kühberger (1998). Thus, for example, even if most participants preferred the sure (or risky) option in *both* framing conditions of the ADP, as long as it were preferred by a significantly *larger* majority in the positively-worded (or negatively-worded) condition, this would suffice for a framing effect to be present.¹⁰

5. Editing, Representation, and Evaluation

Tversky and Kahneman extend prospect theory to explain the pattern of responses to the ADP (and similar risky-choice framing scenarios). They propose that, in the 'be saved' condition, the prospects *sound like* gains, and in the 'die' condition they *sound like* losses. As a result, when the prospect is framed in terms of the number of people who will 'be saved', participants are thought to represent this as a *gain* (presumably relative to the expectation that 600 people will

⁸ I return to the question of precisely what makes wording 'positive' or 'negative' below.

⁹ The distinction between reflection effects and framing effects is also emphasised by Fagley (1993) and Kühberger et al. (1999).

¹⁰ The distinction between reversals and shifts is also discussed by Maule (1989) and Wang (1996) (who calls these 'bidirectional' and 'unidirectional' framing effects).

be killed). Conversely, when the prospect is framed in terms of the number of people who will ‘die’, participants are thought to represent it as a *loss* (perhaps relative to the most desirable outcome of all 600 people remaining alive).

In order to capture these kinds of effects, Kahneman and Tversky (1979, pp. 273-274) posit an initial ‘editing’ phase in decision-making. This involves decision-makers mentally representing the prospects in particular ways – ways which can affect the values they assign to them. Specifically, positively-worded prospects are represented (or ‘coded’) as gains. They are therefore assigned positive values. Conversely, negatively-worded prospects are coded as losses, and are assigned negative values.¹¹

In a subsequent ‘evaluation’ phase, the decision-maker assesses the prospects, based on the values assigned to them. Where prospects have been coded as gains and assigned positive values, decision-makers will tend to be risk-averse. In contrast, where they have been coded as losses and assigned negative values, decision-makers will tend to be risk-seeking. This can explain the observed pattern of responses to the ADP: in the ‘be saved’ condition, where participants treat the prospects as gains, there is a shift towards the sure option. In the ‘die’ condition, where participants treat the prospects as losses, there is a shift towards the risky option.

As is clear from the summary above, it is essential to the prospect theoretic account of framing effects that the use of positive or negative wording cues audiences to represent prospects as gains or losses. However, there is limited discussion in the literature of what counts as positive or negative wording.¹² In what follows, I will think of the positivity or negativity of linguistic expressions as depending straightforwardly on the typical evaluative valence of what the expressions denote. Thus, for example, *being saved* is typically considered good, other things being equal, whereas *dying* is typically considered bad.

Several theorists have noted that risky-choice framing paradigms like the ADP conflate two factors: first, whether the positively- or negatively-valenced predicate expression is used (for example, ‘be saved’ or ‘die’); and, second, whether it is used affirmatively or under negation. For example, the positive wording of ‘be saved’ can be used affirmatively to describe the number of people who *will* be saved, or under negation, to describe the number of people who will *not* be saved. In affirmative contexts, the valence of the overall predicate is reinforced; in negational

¹¹ Kahneman and Tversky discuss various other operations that may occur during the editing phase: for example, probabilities may be combined in order to simplify a problem, and shared components of the prospects may be subtracted out. However, these operations are less relevant for my concerns and I will not discuss them further here.

¹² Indeed, there is a risk of confusing evaluative polarity with something like ‘markedness’, as will be discussed a little further in chapter 7.

contexts, the valence of the overall predicate becomes ‘mixed’, potentially weakening the framing effect.¹³ Levin et al. (1998, pp. 153-158) make similar remarks about confounding or conflicting ‘hedonic tones’ within a single framing condition. While mixed-valence cases constitute an interesting topic for future research, my focus here will be on the more straightforward and clear-cut affirmative uses. It is hoped that a better understanding of these will provide the foundation for further investigation of more linguistically complex structures.

6. (Ir)rationality

An upshot of the prospect-theoretic explanation of framing effects is that human decision-making behaviour is doubly divergent from EUT’s prescription for rational agents. First, in the editing phase described above, prospects are coded as gains or losses purely in virtue of their being described in positive or negative terms, rather than due to their actual effects on overall utility. Standardly, this representational effect is taken to violate the normative principle of description invariance, which specifies that rational agents will respond in the same way to the same outcomes, regardless of how they are described.¹⁴ This first point of divergence from the rational model is peculiar to framing effects. The second point of divergence is common to both framing and reflection effects: in the evaluation phase, decision-makers are risk-averse or risk-seeking depending on whether prospects have been coded as gains or losses. In reflection paradigms, the coding of the prospects corresponds to actual changes in utility; in framing paradigms it does not, since the change in utility is supposed to be identical under each condition.

According to prospect theory, then, people generally fall short of the rational ideal embodied by EUT. Psychologists working in the prospect theoretic tradition have treated this rational failure

¹³ There is some empirical evidence to suggest that risky-choice framing effects may depend on predicate expressions being used affirmatively in the sure options. Kühberger (1995) presents data from two experiments which show that the standard framing effect disappears when the sure options include explicit information about the number of people who will *not* be saved, or will *not* die. Moreover, he finds that the standard framing effect is reversed when the sure options *only* include information about the numbers who will *not* be saved, or will *not* die: in these manipulations, participants were risk-seeking in the ‘be saved’ framing condition, and risk-averse in the ‘die’ framing condition. The findings suggest that it is not merely the use of the positively- or negatively-valenced predicate expression that makes an outcome seem like a gain or a loss but its being used affirmatively rather than under negation. Building on Kühberger’s analysis, Mandel (2001) investigates additional combinations of affirmative and negated uses of ‘be saved’ and ‘die’, across both the sure and risky options of a risky-choice scenario. His results support the hypothesis that framing effects depend on root expressions being used affirmatively (a hypothesis which is further developed and by Tombu and Mandel (2015)). These studies also seem consistent with an earlier set of results published by Van Schie and Van Der Pligt (1995), although those authors describe their findings rather differently.

¹⁴ Tversky and Kahneman (1986, p. 253) describe this principle as “so basic that it is tacitly assumed in the characterization of options rather than explicitly stated as a testable axiom”. In chapter 9 I will develop a refined version of it, which applies more plausibly to agents with limited knowledge.

as a deep fact about human cognition, not merely the result of our limited knowledge, memory, and computational abilities. For example Tversky and Kahneman (1986, p. 252) write:

We argue that the deviations of actual behavior from the normative model are too widespread to be ignored, too systematic to be dismissed as random error, and too fundamental to be accommodated by relaxing the normative system.

Some further evidence that framing effects are counter-normative comes from the fact that alternative framing conditions are typically understood to be equivalent once both conditions are made available.¹⁵ *Implicit* evidence is obtained from within-subjects experimental designs, in which the same participants receive both framing conditions (usually not simultaneously but interspersed with other tasks). Typically, participants give consistent answers once they have the opportunity to see both of the alternative frames, not just one. In other words, participants tend to choose either the sure option or the risky option under *both* frames. Some studies which obtain this result with respect to the ADP, or similar scenarios, include (Frisch, 1993; Kühberger, 1995; Stanovich & West, 1998).¹⁶

Other experiments have elicited *explicit* evidence that people understand alternative framing conditions to be equivalent. They have done so by presenting both frames simultaneously (often after participants have completed a choice task based on one or both frames) and asking participants whether they think the frames are relevantly different. Frisch (1993) finds that most people who responded differently to the two framing conditions of ADP in a within-subjects design nevertheless said that both pairs of prospects should be treated the same, once they had a chance to view them side by side.¹⁷ The within-subjects data suggest that people are *capable* of understanding the alternative framing conditions to be equivalent. And it is standardly held that it would be *rational* for them to do so. On that view, forming distinct representations under each framing condition seems clearly irrational.¹⁸

¹⁵ In chapter 9, however, I will question the status of this evidence.

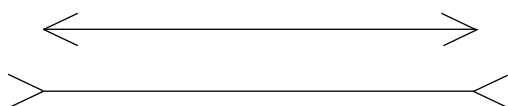
¹⁶ However, there is typically an effect of the *order* in which framing conditions are presented: participants receiving the positively-worded frame first tend to choose the sure option under both frames, whereas those receiving the negatively-worded frame first tend to choose the risky option under both frames. In other words, when first presented with one or other framing condition, participants' responses typically pattern according to the standard framing effect. However, when they view the second framing condition, they tend to treat it as relevantly similar to the first one. That said, substantial minorities of participants still exhibit the standard framing effect even in within-subjects designs (around 30% in both Frisch's and Stanovich and West's ADP-based experiments). And Kahneman and Tversky describe experiments in which a *majority* do: they report that "[the failure of invariance] is not eliminated even when the same respondents answer both questions within a few minutes" (Kahneman & Tversky, 1984, p. 343).

¹⁷ The remainder maintained that the framing conditions should be treated differently and gave various reasons for this. As Frisch notes, though, the experimental design makes it possible that they were merely trying to justify their earlier preference reversal.

¹⁸ The thesis will build the case against this standard view, challenging it directly in chapter 9.

7. Investigating Representational Effects

My research will address the question of how framing affects the representation of prospects in the initial ‘editing’ phase (although the conclusions will also feed through to a different interpretation of the responses produced during the subsequent ‘evaluation’ phase). Prospect theory has little to say about *how* and *why* differences in wording might induce different mental representations. Indeed, Tversky and Kahneman (1981) seem to treat the effects of framing on representation as a brute fact of imperfect human decision-making. They compare such effects to cases of perceptual illusion like the Müller-Lyer paradox (in which two lines of equal length appear different because of the direction of the arrows – see below); or the relative heights of two mountains appearing to change when viewed from different angles (Tversky & Kahneman, 1981, p. 453).



Accordingly, Tversky and Kahneman make the following observation about their experiments using the ADP:

On several occasions we presented both versions to the same respondents and discussed with them the inconsistent preferences evoked by the two frames. Many respondents expressed a wish to remain risk averse in the “lives saved” version and risk seeking in the “lives lost” version, although they also expressed a wish for their answers to be consistent. (Tversky & Kahneman, 1986, p. 260)

As in the perceptual illusions mentioned above, surface cues in linguistic form are conceptualised as pulling participants towards representations of reality which they know to be inconsistent. However, little more is said to develop the analogy into a concrete explanation of how frames are interpreted. The researchers are aware of this gap in their account, noting:

Because the framing of decisions depends on the language of presentation, on the context of choice, and on the nature of the display, our treatment of the process is necessarily informal and incomplete. (Tversky & Kahneman, 1986, p. 273)

The question of how frames are represented has come into sharper focus over the last fifteen to twenty years, as Teigen (2016) points out in surveying developments in quantitative framing research. Nevertheless, it remains understudied. The aim of this thesis is to contribute to filling this gap in our understanding of framing. Before outlining the shape of my argument, I continue tracing the evolution of the framing literature, in order to introduce the distinct paradigm of ‘attribute framing’ which will feature in later chapters

8. Attribute Framing

A survey of the framing literature by Levin et al. (1998) rounds up the first two decades of framing research. The authors helpfully distinguish three kinds of framing paradigm. Alongside the ‘risky-choice’ framing discussed above, their typology identifies ‘attribute framing’ and ‘goal framing’. My research will address only the first two of these; I will set goal framing aside.¹⁹

Levin et al. describe attribute framing as follows:

Attribute framing represents perhaps the simplest case of framing, making it especially useful for gaining a basic understanding of how descriptive valence influences information processing. We label this elementary form of framing “attribute framing” because only a single attribute within any given context is the subject of the framing manipulation. Here, the dependent measure of interest is not choice between independent options, but it is instead a measure of the more basic process of evaluation. (Levin et al., 1998, p. 158)

Below is an example of an attribute framing scenario, taken from an experimental study by Leong, McKenzie, Sher, and Müller-Trede (2017):²⁰

Imagine that you are a recruiter for a college basketball team. Your job is to search for promising high school basketball players and try to recruit them to your college. You are looking through files for players from local high schools, and you are especially interested in players who can score many points.

The file you are currently looking at shows a player whose performance is quite unusual.

Participants in one condition were presented with the following target sentence:

This player made 40% of his shots last season.

¹⁹ Goal framing involves describing an act in terms of the positive or negative consequences of performing it or not performing it. For example, information about breast self-examination (BSE) could be presented in either of the following ways, taken from (Meyerowitz & Chaiken, 1987):

- (i) Research shows that women who do BSE have an increased chance of finding a tumor in the early, more treatable stages of the disease.
- (ii) Research shows that women who do not do BSE have a decreased chance of finding a tumor in the early, more treatable stages of the disease.

The second, negatively worded, frame is found to be more effective than the first in encouraging women to do BSE. A similar pattern is observed in a handful of other studies. However, this framing paradigm introduces a number of complexities, including the use of negation in one frame but not the other. In general, goal framing experiments constitute a relatively heterogeneous class and are not as well studied or as well understood as the other two categories.

²⁰ It is adapted from an experimental scenario discussed by Levin et al. (1998).

Participants in a separate condition were presented with the following alternative sentence:

This player missed 60% of his shots last season.

Participants in both conditions were then asked: “How valuable do you think this player would be to your basketball team?” They answered by rating the player on a scale ranging from ‘Not at all valuable’ to ‘Extremely valuable’.

In this paradigm, both frames are understood to attribute to the player the same property of having made 40% and missed 60% of his shots. However, they do so by explicitly specifying either the proportion ‘made’ or the proportion ‘missed’. Leong et al. found that the player was judged to be more valuable to the team in the ‘made’ condition than in the ‘missed’ condition. This is in line with a catalogue of other attribute framing effects, whereby entities tend to be evaluated more favourably when they are described using positive wording than when they are described using negative wording. The attribute framing effect, then, is a shift in the evaluation of an entity due to its being described using alternative frames.²¹

Similarly to the standard prospect theoretic account of risky-choice framing effects, it has been hypothesised that recipients of attribute frames tend to represent an entity positively or negatively in line with the valence of the words used to describe it (Levin & Gaeth, 1988; Levin et al., 1998). Again, though, theories about how and why the difference in wording affects mental representations are still in their infancy.²²

What is common between risky-choice and attribute framing effects, then, is the idea that they depend on audiences forming different representations of prospects or entities, due to these being framed in different ways. The thesis will explore and develop this basic insight.²³ I will consider exactly how the information in alternative framing conditions is represented. I will argue that, in explaining *why* frames produce different representations, it becomes possible to *rationalise* framing effects.²⁴

²¹ In chapter 6 I will refine the definitions of attribute framing and attribute framing effects. My definitions will be more specific about the linguistic structures involved in these paradigms, and will accommodate structurally similar evaluatively neutral cases, where the framing effect does not depend on the alternative frames having opposing, positive or negative, valence.

²² See chapter 6 for further discussion of the two main competitors: the associationist account and the information leakage account.

²³ Therefore, I will not consider the possibility that *non*-representational mechanisms underlie framing effects (or cognition more generally). Instead, I will work within a broadly representational tradition.

²⁴ My research is thus allied with theories that reinterpret the kind of data marshalled by Kahneman and Tversky as being consistent with rational behaviour. For a helpful survey of the wider debate, see (Kühberger, 2002). Of particular relevance are those who have appealed to linguistic pragmatics in providing rationalising explanations, including Hilton (1995) and Schwarz (1996). Interestingly, Kahneman and Tversky themselves acknowledge that “the conversational aspect of judgment studies deserves more careful consideration than it has received in past research, our own included” (Kahneman

9. Thesis Overview

Applying a philosophical perspective to psychological investigations of framing, I will argue that a handful of semantic and pragmatic features can explain why the representation of a prospect or entity depends on the words used to describe it. Working in the opposite direction, I will also use framing effects as an empirical case study to inform ongoing philosophical debates about the semantics-pragmatics divide. Specifically, I will suggest that the framing data support a relatively minimal conception of semantics and a broadly Gricean conception of pragmatics. Turning to look beyond academia, I will end by exploring speakers' and hearers' responsibilities for the problematic effects of framing in public discourse.

The structure of the thesis is as follows: in chapter 2 I will map out a spectrum of approaches to contextual variability in meaning, signposting the ways in which these approaches will inform my subsequent discussions of the semantics and pragmatics of frames. In chapter 3 I will consider a deflationary explanation of framing effects, according to which frames are represented differently because of their inequivalence at a basic semantic level – specifically, because contrasted predicate expressions like 'be saved' and 'die' fail to jointly exhaust the logical space of possibilities. In chapter 4 I will consider whether framing effects could be explained on the basis of number expressions obtaining lower-bounded interpretations, such that, for example, '200' is understood as denoting *at least 200*. These chapters will raise important methodological points, showing the need for greater care in setting up alternative framing conditions. Nevertheless, I will conclude that it is still possible for framing effects to arise even when the alternative frames are genuinely logically equivalent. Before embarking on a different explanation, in chapter 5 I will address a possible objection to the claim that frames *qua* sentences can be logically equivalent to one another. This claim presupposes that sentences can be truth-evaluable, which is prominently denied by proponents of Radical Contextualism. I will defend a liberal understanding of 'truth-evaluability', according to which contents count as truth-evaluable as long as they can map at least one possible world to at least one truth-value. On this understanding, alternative frames would count as truth-evaluable and capable of logical equivalence.

In chapter 6 I will introduce the 'information leakage' account of attribute framing effects, endorsing a version of the view known as the 'reference point hypothesis'. In chapter 7 I will argue that the 'reference point information' typically conveyed by frames can be analysed as a Gricean conversational implicature. In chapter 8 I will suggest that there is little plausibility or

& Tversky, 1982, p. 135). However, this point is often lost in the standard interpretation of framing effects as being evidence of irrational behaviour.

parsimony in taking reference point information to 'intrude', in one sense or another, on what is 'said'. Finally, in chapter 9 I will show how my analysis shifts the responsibility for framing effects away from hearers. I argue that they may behave perfectly rationally in forming different representations on the basis of alternative frames. Responsibility will be shifted, instead, towards *speakers*. In this chapter, I seek to apply the theoretical framework developed throughout the thesis to real-world cases of framing. In this way, the philosophical and psychological conclusions of my research will be connected with issues of contemporary social and political relevance.

CHAPTER 2

Semantic Conceptions

Abstract

Within the philosophy of language, there is now a well-established spectrum of views about whether, when, and how the meanings associated with linguistic utterances depend on extra-linguistic factors. At one end of the spectrum, ‘Semantic Minimalism’ seeks to carve out a set of minimal semantic meanings, which attach to sentences in a language largely (or entirely) independently of how they are used. These meanings are distinguished from the wider pragmatic meanings a sentence is used to convey on each occasion of its use. At the other extreme, ‘Occasion-sensitivity’ treats the meanings of uttered sentences as being richly and essentially dependent on extra-linguistic features of the wider context. In this initial set-up chapter, I sketch how each approach could be brought to bear in explaining framing effects, and I signpost the more detailed discussions coming up in later chapters.

1. Introduction

The philosophical debate with which I wish to engage concerns variations in meaning, broadly construed, which arise when the *same* linguistic expression is used on *different* occasions. After all, expressions can be used at different times, by different speakers, in addressing different audiences, and with a variety of different goals. Changes in these wider extra-linguistic circumstances can cause a given expression to acquire very different interpretations.

Unlike examples of meaning variation that are standardly discussed in the philosophical literature, framing effects arise from the use of *distinct* linguistic expressions, on occasions that are *qualitatively identical* (since, in experimental studies at least, wider contextual factors are controlled for). Nevertheless, there are important parallels between both kinds of case. In particular, alternative frames are still supposed to express the *same* meaning as one another in some sense (the nature of which will be explored throughout the thesis). It is the precise words chosen to convey that meaning which plays the role of an occasion-specific variable here. In later chapters, I will demonstrate how the case study of framing effects sheds new light on philosophical debates about meaning variation, informing our conceptions of semantics, pragmatics, and the boundary between the two. The purpose of this initial chapter, however, is to survey the current state of that debate, thereby laying the groundwork for my contribution to it.

The structure of the chapter is as follows: in §2 I set up the debate by outlining a simplified model of the linguistic interpretation process. I then describe how certain aspects of this process might be characterised by six distinct theories of linguistic meaning and

communication:²⁵ Semantic Minimalism (discussed in §3); Indexicalism (discussed in §4); Conventionalism (discussed in §5); Contextualism (discussed in §6); Semantic Relativism (discussed in §7); and Occasion-sensitivity (discussed in §8). Throughout the discussion, I will consider the resources available to each theory for describing the effects of framing on meaning.

2. Interpretation

Upon encountering a linguistic stimulus, the audience is faced with a stream of perceptual input (phonological, orthographic, or gestural, depending on the format in which the stimulus is presented).²⁶ I will assume that hearers aim to recover the meaning(s) that the speaker intended to communicate (which is nevertheless compatible with their also having a range of other aims). In pursuing this aim, hearers need to transform perceptual inputs into meaningful information. As a first-pass characterisation of this process, perceptual inputs can be thought of as being mapped by hearers to certain linguistic *types*, i.e. expressions in a natural language like English. Thus, for example, the following set of marks corresponds to the English word, ‘survive’:

s u r v i v e

And the following (partly overlapping) set of marks is mappable to the English sentence ‘John survives’:

J o h n s u r v i v e s .

Expressions generated via this process can then be thought of as being mapped to their *conventional meanings* in the language. Conventional meanings are understood here as lexical entries for linguistic expressions, stored in a hearer’s ‘mental lexicon’ – a kind of memorised database of such entries. Thus, for example, the English expression, ‘survive’ is mappable to the concept *survive*. Concepts – however exactly they should be characterised – can be thought of as having certain denotations. For example, the concept *survive* would denote a set of entities that survive (where this set might be relativised to possible worlds, times, and perhaps other parameters).

Where the linguistic stimulus is a complex expression, containing multiple simple expressions, the meanings of its constituent expressions will need to be combined in accordance with the

²⁵ Five of these are drawn from the typology developed by Borg (2012). I add a sixth – ‘Conventionalism’ – which has been gaining in prominence in recent years.

²⁶ For ease of exposition, I will use ‘utterance’ as a catch-all term for any use of a linguistic stimulus, and ‘speaker’ and ‘hearer’ for its producer and audience, regardless of the format of presentation.

compositional rules for the language.²⁷ For example, the sentence ‘John survives’ is mappable to a representation that we might write as *survive(John)*, where the concept *survive* takes *John* as its argument. Standardly, the mental representations to which sentences ultimately get mapped are held to be *truth-conditional*: thus, for example, *survive(John)* should output *TRUE* if John is a member of the set of entities that survives, and *FALSE* if he is not (and, perhaps, some third truth-value if it is indeterminate whether or not he survives).

This simplistic reconstruction of the interpretation process is immediately complicated, however, by cases in which the representations hearers form are clearly not – and cannot be – merely the result of combining the conventional meanings of constituent expressions but depend instead on extra-linguistic factors.²⁸ Taking the exchange below as an example, imagine that A and B are discussing an unseeded pair of tennis players, Kris and Marta, who are about to face the world number ones in the first round of the Wimbledon doubles tournament.

A. There’s a lot of hype surrounding the match this afternoon.

B. They won’t survive.

A’s interpretation of B’s utterance of (1) is likely to be affected by the wider discourse context.

(1) They won’t survive.

First, the pronoun ‘they’ is an ‘indexical’ expression, meaning that its denotation varies across different occasions of its use. In our imagined scenario it obviously refers to the unseeded pair of players, Kris and Marta. However, this fact is not determined purely by the conventional meaning of ‘they’. Instead, it is determined partly by the fact that A and B have been discussing those players.²⁹ Second, the expression ‘survive’ concerns the players’ progress through the tournament, not their continued existence as living beings. This suggests that ‘survive’ may need to be mapped to a concept that is more specific, or different, than its conventional meaning, in a process that is known as ‘modulation’ of the conventional meaning. Third, presumably what is meant by the utterance of (1) in this scenario is that the players will not progress through *to the next round of* the tournament. This is despite the fact that (1) contains

²⁷ As we will see below, views differ as to whether composition takes place before or after occasion-specific modulation or enrichment of conventional meanings. In setting out the first-pass account of interpretation, I set that issue aside; however, I will begin to problematise this overly simplistic account immediately below.

²⁸ Any effects on interpretation beyond those of conventional meanings and compositional rules I will label ‘extra-linguistic’, even though they will sometimes relate to other linguistics aspects, such as the surrounding discourse.

²⁹ I will remain neutral as to whether the denotations of indexical expressions like ‘they’ are constitutively determined by the speaker’s intentions or by purely objective features of the utterance situation. This issue is the subject of ongoing debate. For a prominent intentionalist account, see Neale (2016); for the opposing objectivist view, see (Gauker, 1998, 2008, 2019).

no corresponding explicit component. That points to the need for some further enrichment of the interpretation, perhaps to include an ‘unarticulated constituent’ (Perry, 1986).³⁰ Overall, then, a hearer of (1) might form the following representation:³¹

(1a) Kris and Marta won’t progress to the next round of the tournament.

The utterance of (1) might also bring various other informational contents to mind, for example that the supporters of the unseeded players will be disappointed. Again, such representations cannot be derived merely by composing the conventional meanings of (1)’s surface constituents. Instead, they depend in large part on hearers making inferences on the basis of their wider knowledge (for example, by inferring that the hype will have led the supporters to raise their expectations, and recalling that supporters are usually disappointed when their team loses). It is generally agreed that wider representations of this kind are pragmatic ‘implicatures’, along the lines discussed by Grice (1989). I will describe the defining features of implicatures in much more detail in chapter 7. For now, I will proceed with an intuitive grasp of the distinction between relatively indirect implicated meanings and meanings which seem to be expressed more directly, like the one captured by (1a).

It is relatively uncontroversial that extra-linguistic factors will affect hearers’ *interpretations* of linguistic stimuli. Far more contentious is the question of how far anything characterisable as the *meaning* of an utterance might depend on these factors. Some theorists have sought to isolate *minimal semantic meanings* that remain (largely or entirely) independent of extra-linguistic factors. Others have rejected any appeal to such entities, arguing that meanings inevitably depend on wider contextual phenomena. This is the crux of the debate between Minimalism and Contextualism, which I survey throughout the next six sections.

3. Semantic Minimalism

According to one family of theories, every component of a complex linguistic expression’s semantic meaning must be traceable back to the conventional meanings of its constituents or the compositional rules that govern their combination. Theories in this camp bifurcate into Semantic Minimalism and Indexicalism. In this section I will discuss Semantic Minimalism (also known as ‘Insensitive Semantics’, ‘Invariant Semantics’ or, simply, ‘Minimalism’). As we will see,

³⁰ Borg (2016, pp. 341-342) proposes that modulation and the addition of unarticulated constituents may be, in principle, interchangeable.

³¹ Of course, the precise structure of a language user’s mental representation of (1) may turn out to be far less like a sentence of English than (1a) is; however, I will continue to work with this kind of intuitive formulation for ease of exposition.

Minimalists recognise only a relatively small set of context-sensitive elements in language. In §4 I will discuss how Indexicalists have been moved to enlarge that set.

3.1. Context-sensitivity

Minimalists standardly recognise a set of context-sensitive expressions, the denotations of which vary across different occasions of use. The set of context-sensitive expressions includes ‘indexicals’ (like ‘I’) and ‘demonstratives’ (like ‘this’ and ‘that’). Such expressions must be paired – through a process known as ‘saturation’ – with suitable contextual parameters, such as particular individuals (as in ‘I’ or ‘you’) or objects (as in ‘this’ or ‘that’).³² Returning to sentence (1), Minimalists would standardly accept that the English expression ‘they’ is context-sensitive.

Different versions of Minimalism have proposed different criteria for inclusion in the set of context-sensitive elements of language. Cappelen and Lepore (2005) describe a ‘Basic Set’ of such elements, which is more or less exhausted by those enumerated by Kaplan (1989): personal pronouns, demonstrative pronouns, syntactic markers of tense, and a small number of open-class words. In contrast, Borg (2012) allows that the set of context-sensitive elements could potentially be enlarged, so long as there are formal semantic or syntactic justifications for adding further members. It is expected, though, that the linguistic evidence will continue to support the presence of relatively limited context-sensitivity in language.

Minimalists typically argue that genuinely context-sensitive expressions like ‘they’ demand an extra-linguistic contribution to the semantic meaning of an utterance like (1).³³ This is because the influence of extra-linguistic factors can be traced back to the expression’s conventional meaning. Thus Borg writes:

[W]hat minimalism should be held to claim is that the route to semantics runs along exclusively lexical and syntactic footholds (Borg, 2012, p. 206).

We might therefore represent the minimal semantic meaning of (1) as follows:

(1b) Kris and Marta won’t survive.

³² Although I use the terminology of ‘context-sensitivity’, I intend to remain neutral about whether the extra-linguistic factors that constitutively determine denotations are speakers’ intentions or features of a (formal or situational) context, as noted in footnote 29. I will also defer questions about the extent to which conventional meanings constrain the role of extra-linguistic factors. For example, the conventional meaning of ‘I’ might be thought to require, quite specifically, that the referent be the speaker (although see (Predelli, 1998) for discussion of some complicating factors). In contrast, the conventional meanings of the demonstratives ‘this’ and ‘that’ may seem to provide far looser constraints on the referent. Since this issue is orthogonal to my project, I will simply set it aside for now.

³³ Such meanings correspond to the ‘contents’ posited by formal semanticists working in the Kaplanian tradition.

Minimalists are also united in *denying* that semantics should capture *modulated* meanings, which depend on wider aspects of the discourse situation (such as the modulation of ‘survive’ to mean something like *progress*). Neither will minimal semantic meanings incorporate any *unarticulated constituents* (as in the inclusion of ‘*to the next round of the tournament*’ in (1a)). And, of course, Minimalists exclude from semantics any wider implicatures that hearers may derive from a speaker’s utterance. In sum, extra-linguistic effects on semantic meaning are limited just to those which are required by constant, fixed features of language.³⁴

3.2. Propositionalism

Another distinction between the cluster of views in the Minimalist camp concerns ‘propositionalism’, the claim that all well-formed declarative sentences express complete, truth-evaluable, propositions. Borg (2004, 2012) argues that they do (at least, once sentences containing context-sensitive elements have been suitably relativized to extra-linguistic facts). In contrast, Bach (2001, 2006a) advocates a ‘radical’ form of Minimalism, according to which semantic meanings will sometimes be incomplete. He writes:

If the semantic content of a sentence is capable of being true or false, is a possible content of thought, and is the possible content of an assertion, then the sentence expresses a proposition, otherwise only a propositional radical. (Bach, 2006a, p. 440)³⁵

Bach’s view is sometimes categorised as Contextualist, rather than Minimalist, on the grounds of this incompleteness claim (Borg, 2012; Borg & Fisher, forthcoming). However, to the extent that Bach believes ‘propositional radicals’ play an important theoretical role, that would set him apart from other Contextualists, who generally consider such entities to be theoretically inert (as will be discussed in §6 below).

3.3. Theoretical Role

On one hand, Minimalists acknowledge that minimal semantic meanings need not correspond to what ordinary language-users judge to have been communicated (including whether the speaker has said something true or false). For example, they may accept that (1) intuitively

³⁴ Borg (2012) calls these ‘lexico-syntactic’ features; I will sometimes refer to them as ‘linguistic’ here. In later chapters I will want to distinguish linguistic effects that depend on conventional *meaning* from those which depend on properties of *use*.

³⁵ Cappelen and Lepore are harder to pin down on this issue. In their ‘Reply to Bach’ they avoid committing themselves to propositionalism (Cappelen & Lepore, 2006). However, they clearly do take many sentences to have complete, truth-evaluable semantic meanings, where non-Minimalists would typically consider them to be underdetermined.

communicates something like (1a). However, they deny that this must be captured in full by the minimal semantic meaning. In general, as Borg writes:

[M]inimalists are happy to reject the idea that a semantic theory should limn our intuitive judgements of what is said by the utterance of a sentence in a given context (Borg, 2012, p. 49).

Nevertheless, it is maintained that minimal semantic meanings play *some* important role in theorising. This has been cashed out in a handful of different ways. Cappelen and Lepore (2005) have argued that minimal semantic meanings are invariably *part of* what is communicated. They write:

In short: the proposition semantically expressed is our minimal defense against confusion, misunderstanding, mistakes and it is that which guarantees communication across contexts of utterances. (Cappelen & Lepore, 2005, p. 185)

They go on to say:

If there's a difference between having a cognitive function and corresponding to a stage in processing or having a psychological reality, we don't know what that difference consists in. (Cappelen & Lepore, 2005, p. 186)

Borg (2004, 2012) denies that minimal semantic meanings need be psychologically realised in the way Cappelen and Lepore suggest. Instead, she allows, for example, that a hearer of (1) might represent (1a) without first deriving (1b) in full. According to Borg, there is a mental module which has the teleological function of generating minimal semantic meanings during comprehension. Following Fodor (1983), Borg proposes that linguistic processing is *modular* (see, especially, chapter 2 in (Borg, 2004)). In other words, it is undertaken by a discrete cognitive system, operating independently of – despite being situated within – a wider set of predominantly non-modular thinking and reasoning processes. A mental module is understood here as “an encapsulated body of information with deductive, computational rules operating only over that information” (Borg, 2004, p. 78). The language processing module operates over *linguistic* inputs (simple or complex expressions of language) and transforms them in accordance with classical computational operations into further representations. Importantly, Borg departs from Fodor in arguing that the module can provide truth-conditional semantic meanings of natural language sentences, rather than being restricted to syntactic and phonetic processes. Borg maintains that a linguistic input “can be processed, up to and including semantic processing, without appeal to information outside the language faculty” (Borg, 2004, p.

141). In other words, it is claimed that a world-involving representation is derivable in the absence of any extra-linguistic information.³⁶

Borg argues that a hearer will derive the minimal semantic meaning of an utterance just in case the semantic module completes its processing on that occasion. That will depend on two conditions obtaining: first, the hearer must possess “adequate lexical resources” to understand the conventional meanings of the constituent expressions (Borg, 2012, p. 63). Second, it is necessary that “attentional resources are not diverted from processing literal meaning” (ibid). This ensures that semantic processing unfolds in accordance with the hearer’s underlying linguistic competence. In sum, although the semantic module has the function of transforming utterances into minimal semantic meanings, its doing so on any given occasion is defeasible. In some instances, no minimal semantic meaning is psychologically realised and will not, therefore, be consciously accessible. This distinguishes Borg’s minimal semantic meanings from those of Cappelen and Lepore (and also from Recanati’s notion of ‘what is said’, as we will see in §6). Nevertheless, the counterfactual claim remains: a minimal semantic meaning would have been derived under suitable conditions. This is taken as a psychological justification for positing minimal semantic meanings.

Harris (forthcoming) presents a similarly modular analysis of semantic processing. Like Borg, he grounds semantics in the operations and representations of an informationally encapsulated psychological system. However, Harris allows that the outputs of semantic processing may be incomplete, non-truth-evaluable, sub-propositional representations. In particular, he does not believe that the denotations of context-sensitive elements of language contribute to semantic meanings; instead, semantics merely involves the composition of invariant conventional meanings.³⁷ Although the resulting picture shares some features with Contextualist accounts, Harris assigns an important psychological role to the minimal semantic meanings he posits. Indeed, seen from a certain perspective, his view of semantics is even more minimal than that of most Minimalists: it requires not only that semantic meanings be *traceable back to* conventional meanings and composition rules but that they *depend solely on* these.

As well as their having reality in individual psychology, minimal semantic meanings have been argued to play other roles, including in our interpersonal communicative practices. For

³⁶ This point is, at least *prima facie*, in tension with the claim that context-sensitive elements in language contribute occasion-specific, extra-linguistically determined denotations to minimal semantic meanings. Borg (2004, 2012) seeks to resolve any apparent tension by arguing that the denotations of context-sensitive elements may contribute to semantic meanings without the hearer being able to recognise what the denotations are. This response has been challenged, for example by Harris (forthcoming). Since the dispute is not directly relevant to my argument, I will not attempt to resolve it here.

³⁷ This avoids the potential tension in Borg’s account, discussed in the previous footnote.

example, as is evident from the earlier quote, Cappelen and Lepore (2005) see them as the meanings interlocutors fall back on when they hit obstacles in communicating with one another. More recently, Borg has argued that minimal semantic meanings have practical import in communicative contexts (such as legal contexts) which require speakers to assume ‘strict liability’ for their words (Borg, 2017; Borg & Connolly, forthcoming).

3.4. Pragmatic Meaning

Minimalists readily acknowledge the importance of contents like (1a), which result from further modulation or enrichment of the minimal semantic meaning. However, they consider these to be *pragmatic* meanings, which concern what speakers *do* when they use words – the communicative effects they aim to bring about.³⁸ Thus, as Borg puts it, “minimalism drives a wedge between literal sentence meaning and pragmatic speaker meaning” (Borg, 2012, p. 215). She characterises the distinction as follows:

[S]emantic content is one kind of thing (a repeatable, codifiable, rule-governed kind of thing) while speech act content is another kind of thing altogether (a potentially unrepeatable, nebulous, context-governed kind of thing) (Borg, 2012, p. 15).

In particular, Borg (2004, 2012) links pragmatics to non-modular processes of central cognition, which can potentially draw on the full range of information available to an interpreter.³⁹ The sharp distinction between semantics and pragmatics is complemented by an important continuity between different pragmatic processes and meanings. That said, it remains open to Minimalists to further differentiate among varieties of pragmatic effects. Indeed, Borg (2017) proposes that speakers can be ‘conversationally liable’ for some (asserted) pragmatic contents but not others (which are merely implicated).

3.5. Minimalism and Framing

A Minimalist approach to explaining framing effects could proceed in two ways: first, the alternative frames might be argued to express inequivalent minimal semantic meanings; and

³⁸ The literature is not always clear about whether pragmatic meanings should be thought of as what the speaker intended to communicate, what the hearer understood, or both. I will remain neutral on this point, although see chapter 7 for further discussion.

³⁹ Relatedly, Borg maintains that whenever the conventional meaning of a context-sensitive expression specifies that the denotation depends on the speaker’s intended referent, the hearer’s ability to recognise that referent will be the result of *pragmatic* interpretation, not *semantic* processing. The idea is that reasoning about others’ mental states paradigmatically involves competences that are not language-specific but integrate wider information. Cappelen and Lepore (2005), who do not link their Minimalist account to a modular cognitive mechanism, are more flexible on this point: they allow that semantic interpretation may involve hearers reasoning about extra-linguistic information of both objective and subjective kinds.

this could explain why each sentence gets represented differently.⁴⁰ Alternatively, the Minimalist might argue that alternative frames have equivalent minimal semantic meanings but are represented differently due to their being pragmatically enriched in different ways.

In chapter 3 I will explore a proposal that is consistent with the first approach. I consider there whether framing effects could arise from features of the conventional meanings of predicate expressions substituted across framing conditions (like ‘be saved’ and ‘die’) – specifically, their failure to jointly exhaust the logical space. The discussion in chapter 4 could also be cast in a Minimalist light: I discuss there the extent to which framing effects are driven by lower-bounded representations of number expressions (so that, for example, ‘200’ is represented as denoting *at least 200*). For those who consider number expressions to be *semantically* lower-bounded, this effect would be traceable back to conventional meaning.

In chapters 6-8 I will consider a hypothesis that falls within the second, pragmatic approach. This predicts that the use of one or other frame typically results in its minimal semantic meaning being pragmatically enriched with ‘reference point information’. In fact, in chapter 8 I will argue that the framing data provide some independent support for Minimalism. For now, however, I proceed to outline five competing approaches to Minimalism.

4. Indexicalism

In one way or another, all non-Minimalists argue that the effects of extra-linguistic features are more pervasive and penetrating than Minimalists allow. A popular strategy for trying to show this is by appeal to what Cappelen and Lepore (2005) label ‘context shifting arguments’. As Hansen and Chemla (2013) point out, context shifting arguments (CSAs) in fact include two distinct steps:

It is helpful to think of a context shifting argument as consisting of two parts: (i) a *context shifting experiment*, which elicits intuitions about uses of an expression *e* in different imagined contexts, and (ii) an argument that the best way to explain the intuitions generated in response to the experiment involves semantic features of *e*.
(Hansen & Chemla, 2013, p. 287)

In the first step, then, context shifting experiments aim to show that, when a sentence is evaluated for truth against a fixed state of affairs, it may be judged true in some contexts and false in others. Take, for example, the following sentence:

⁴⁰ The explanation would need to be combined with an explanation of why people typically judge the frames to be equivalent when both are presented. Presumably, that would involve their being further pragmatically enriched in such contexts. My focus in the thesis will be primarily on separate presentation contexts, although I briefly discuss joint presentation contexts in chapter 9.

(2) John was old when he died.

Imagine that (2) is used of King John of England, who died at the age of 49 in the year 1216. In many contemporary discourse settings, an utterance of (2) would be judged to express something false: after all, current life expectancy for English men is around 80 years. Nevertheless, it is possible to imagine situations in which an utterance of (2) intuitively states something true. For example, imagine that the interlocutors are discussing average male life expectancy in the middle ages (which was far below 49 years) and the relative longevity of those born into royalty. An utterance of (2) in such a setting might intuitively be judged to express something true. Note that, in both scenarios, the age at which John died remains fixed; it is only the change in the circumstances in which (2) is uttered that results in the shift in its truth-value.

In the second step of the CSA strategy, non-Minimalists argue that sentences like (2) must fail to express constant semantic meanings; and they argue that many sentences are like (2) in this respect.⁴¹ Moved by the CSA data, a group of theorists whom Borg (2007, 2012) describes as 'Indexicalists' have posited a far larger set of context-sensitive elements in language than Minimalists standardly allow. I follow Borg in including under the banner of Indexicalism a cluster of overt and covert approaches, which are discussed briefly below.

Advocates of an *overtly* Indexicalist stance include Szabo (2001) and Rothschild and Segal (2009). They argue for large amounts of hitherto unnoticed indexicality in language, claiming, for example, that many ordinary predicate expressions specify a role for extra-linguistic factors in fixing their denotations. These expressions include colour terms like 'red' and, possibly, gradable adjectives like 'old'. It is held that such expressions function similarly to 'I', 'you' and 'here', having (sometimes multidimensionally) context-sensitive lexical entries.

A different, *covertly* Indexicalist account has been put forward by Stanley and others (King & Stanley, 2005; Stanley, 2002, 2005; Stanley & Szabo, 2000). These theorists argue that the underlying logical forms of sentences contain covert variables, which call for extra-linguistically determined values. Stanley writes:

On this approach, we may discover that the 'real' structure of a natural language sentence is in fact quite distinct from its surface grammatical form. Talk of logical form in this sense involves attributing hidden complexity to sentences of natural language, complexity which is ultimately revealed by empirical inquiry. It is in this sense that I

⁴¹ In particular, Travis (2006, 2008) provides a wealth of examples to show how ubiquitously truth-value judgements can shift across contexts.

intend the thesis that all context-dependence is traceable to logical form. (Stanley, 2000, p. 392)

Stanley posits a variety of covert variables. Sentence (2), for example, is analysed as having something like the following underlying structure:

(2a) John was old(x) when he died.

In (2a) the variable 'x' is a placeholder for a class of entities against whom John is implicitly being compared. Returning to our earlier scenario in which 'John' refers to King John of England, in many contexts the comparison class will be something like *English males alive at the time of utterance*. The semantic meaning of (2), as uttered in 2020, can then be represented roughly as follows:

(2b) John was old when he died, relative to English males alive in 2020.

Clearly, (2b) is false. However, we can also imagine discourse situations in which the value of x is set to something different, such as *English males living in the middle ages*. That would generate the following (true) representation:

(2c) John was old when he died, relative to English males living in the middle ages.

The suggestion, then, is that the hidden comparison class variable can take on different values on different occasions, accounting for the way in which truth-value judgements shift across utterances of (2).

Similarly, Stanley has sought to deal with cases of 'domain restriction' by arguing for covert domain variables. For example, an utterance of 'Every bottle is in the fridge' may naturally be understood as concerning only a small subset of all the bottles in the universe. To capture that fact, Stanley analyses the logical form of the sentence as 'Every bottle(x) is in the fridge', where 'x' acts as a placeholder for some contextually relevant set of bottles.⁴² Elsewhere, Stanley appeals to covert location variables in weather reports like 'It is raining' to capture the intuition that such reports concern particular locations, rather than just any possible location.

In general, then, the Indexicalist strategy seeks to trace semantic meanings back to constant linguistic features at the same time as aligning them with our intuitive judgements about the meanings and truth-conditions of utterances. As Stanley and Szabó put it:

⁴² In fact, the variable Stanley proposes is a more complex function, which can map an object to a property (Stanley, 2005; Stanley & Szabo, 2000). However, the technical details of his analysis are not important for present purposes.

[A]ccounting for our ordinary judgments about the truth-conditions of various sentences is the central aim of semantics (Stanley & Szabo, 2000, p. 240).

This motivation is precisely what the Minimalist objects to, preferring instead to tie semantics *purely* to linguistic rules and conventions. Indeed, the Indexicalist's attempt to do justice to both the linguistic data and intuitive truth-value judgements leads to difficulties which I will discuss briefly in §6 below.

4.1. Indexicalism and Framing

Taking inspiration from Indexicalism, we might be tempted to argue that frames include (overtly or covertly) context-sensitive elements, which can acquire distinct values in use; and perhaps this causes alternative frames to be semantically inequivalent.⁴³ I will consider – but reject – this possibility in chapter 8, when I assess whether the ‘reference point information’ conveyed by alternative frames might intrude on truth-conditional semantic meaning.

5. Conventionalism

I now move on to consider a cluster of views put forward during the last decade, which I group together under the heading of ‘Conventionalism’. Somewhat similarly to Minimalists and Indexicalists, Conventionalists consider semantics to be associated with linguistic rules and conventions. However, these are understood more expansively, to include properties of language *use*.

5.1. Ambiguous Conventions

A prominent account falling under the Conventionalist umbrella is put forward by Lepore and Stone (2015) (and is further developed in (Stojnić, Stone, & Lepore, 2017)). The authors identify a class of ‘interpretive effects’ that are standardly classified as *pragmatic* but which, they argue, in fact depend on linguistic conventions.⁴⁴ It is argued, for example, that the English expression ‘can’ is conventionally used to make requests, as well as to ask about what is possible. So, for example, sentence (3) is held to be ambiguous between *asking* whether John is capable of choosing and *requesting* that John choose.

(3) Can John choose?

⁴³ Presumably, these can also be set to the *same* value in certain situations, to capture instances where the frames are treated as equivalent.

⁴⁴ Lepore and Stone are thinking of conventions here in Lewisian terms, as arbitrarily selected solutions to coordination problems.

Both uses are considered to be determined by the linguistic conventions of English, with extra-linguistic, pragmatic considerations bearing only on the task of selecting which interpretation is appropriate on that occasion (as in standard cases of linguistic ambiguity).

In contrast with mainstream semantic theory, Lepore and Stone jettison the distinction between conventions of *meaning* and conventions of *use*. Instead, they lump both together within a broad category of conventional linguistic effects. They map this category to what gets onto the conversational 'scoreboard' or 'record', which captures the information that is made publicly available to interlocutors during communication. This conventionally-determined information, which is held to coincide with what appears on the conversational record, is treated as 'semantic' in their account.

It seems possible, though, that Lepore and Stone wish to retain a distinction between truth-conditional and non-truth-conditional conventional semantic effects. They write:

Semantics, on our view, can be taken to include all the linguistic information that we use to recover the content that speakers contribute to a conversation through their utterances. We have argued that many different kinds of conventional linguistic information are recruited in the interpretation of an utterance in addition to its grammatical structure and the conventional specification of its truth-conditional content. (Lepore & Stone, 2015, p. 265)

Depending how closely any distinction between truth-conditional and non-truth-conditional effects tracks the distinction between conventions of meaning and conventions of use, the resulting view may turn out to carve up the terrain in substantively the same way as other approaches (the difference being purely terminological). Nevertheless, the way Conventionalists expand the concept of semantics at least reveals an important difference in emphasis: the suggestion is that the important joint to be carved lies between conventional and non-conventional effects, not between conventional *meanings* and other effects on interpretation.

5.2. Biases

Another example of Conventionalism is Predelli's theory of 'bias' (Predelli, 2013). Predelli argues that many expressions have conventional 'biases' in addition to their conventional meanings (or 'characters'). These biases constrain the situations in which the expressions may be used. Summarising the view, Predelli writes:

The conventional meanings of some expressions, so I suggested, exceed their characters, and impose restrictions on the class of context of use which are independent of the peculiarities of this or that type of use. (Predelli, 2013, p. 78)

To take one of his examples, the expression ‘hurray’ can only be felicitously used in situations where (roughly speaking) the speaker has a favourable attitude towards the state of affairs being described. Conversely, ‘alas’ can only be used where the agent has an *unfavourable* attitude.⁴⁵ According to Predelli, then, sentence (4) requires that the speaker be happy that John survived; sentence (5) on the other hand, requires that the speaker regret that fact.

(4) Hurray, John survived.

(5) Alas, John survived.

The biases Predelli posits are not taken to affect the truth-conditions of an uttered sentence. Instead, sentences like (4) and (5) are held to share a single set of truth-conditions.⁴⁶ Nevertheless, their distinct (indeed, opposing) biases result in their having different use-conditions, and this is taken to be part of their semantics.

5.3. Conventionalism and Framing

The Conventionalist might seek to explain framing effects semantically, by appealing to distinct conventions applying to alternative frames.⁴⁷ In chapter 7 I will open up the possibility that alternative frames are conventionally used in different ways. However, in chapter 8 I will argue that, even if they are, this should not be thought of as affecting their *semantics*. Instead, the framing data are more parsimoniously explained by holding apart *meaning* conventions (which contribute to semantics) from *use* properties (which affect what can be pragmatically conveyed).

⁴⁵ In fact, the biases of ‘alas’ and ‘hurray’ are held to exhaust their meaning entirely; such expressions are therefore ‘purely biased’ (Predelli, 2013, p. 67). Note that, in this example, the relevant use-conditions concern the speaker’s attitude. However, in principle, biases could invoke various other conditions.

⁴⁶ In this respect, biases are similar to how some have conceptualised presuppositions, which will be discussed in relation to framing in chapter 8. Predelli argues that biases survive in syntactic environments where presuppositions are blocked, including under attitude operators (Predelli, 2013, p. 75). However, it is worth noting that the logical tests for presupposition are disputed, and their behaviour under attitude operators is particularly controversial (Beaver & Geurts, 2014, §5). Thus, some theorists, like Potts, argue that presuppositions are ‘plugged’ by attitude predicates (Potts, 2015). Others argue that presuppositions can project from such contexts (Geurts, 1998). Therefore, it may ultimately be difficult to hold biases and presuppositions apart, although I will not pursue the point here.

⁴⁷ Presumably, though, there would need to be other conventions that allow the frames to be interpreted as equivalent on some occasions.

6. Contextualism

A challenge for Indexicalists is to demonstrate that the effects they seek to capture semantically are genuinely traceable back to language itself, rather than arising purely from contextual factors. Stanley attempts to meet this challenge by appealing to formal semantic data. He argues that hidden variables must be posited in order to generate certain *bound* readings of the constructions he considers, and he generalises from these to unbound cases. However, this ‘binding principle’ has been widely criticised (Borg, 2012; Cappelen & Lepore, 2002; Collins, 2007; Hall, 2008; Neale, 2007b; Recanati, 2004; Rothschild & Segal, 2009). A similar obstacle looks set to face Lepore and Stone’s Conventionalist position, insofar as they claim that anything affecting the ‘conversational scoreboard’ can be traced back to linguistic conventions.

The difficulties involved in tying intuitive judgements about truth and meaning directly to constant features of language have led many to take a different approach to context shifting arguments. In this section, I will focus on Contextualism. Contextualists allow that the meanings tracked by our intuitive truth-value judgements may be ‘freely’ affected by extra-linguistic factors. Free effects cannot be traced back to conventional meanings or compositional rules, arising instead from wider features of the context.

Moderate forms of Contextualism accept the coherence of positing minimal semantic meanings, at least sometimes. For example, Recanati (2010) takes his ‘Truth Conditional Pragmatics’ (TCP) to be compatible with a form of Minimalism he dubs ‘S*-Minimalism’:

S*-Minimalism is still compatible with TCP; for the level of meaning it posits, which satisfies the minimalist constraint by definition, need not be the same level of meaning as that which concerns TCP, namely the intuitive truth-conditions of the utterance (what is said_{int}); hence there need be no contradiction between TCP’s nonminimalist characterization of what is said_{int} and S*-Minimalism (Recanati, 2010, p. 13).⁴⁸

Nevertheless, Moderate Contextualists deny that minimal semantic meanings are of any theoretical value in their own right, and argue that they are often entirely irrelevant. In particular, minimal semantic meanings are held to play no privileged role in the individual psychologies of speakers and hearers, not being the product of any discrete mental system. Related to this, Contextualists have argued that modulation and enrichment can take place *in between* the retrieval of conventional meanings and the composition process (Recanati, 2010). It is argued that those processes happen alongside the saturation of any context-sensitive

⁴⁸ Carston (2009) similarly allows that sentences may express minimal semantic meanings. However, she is sometimes less ecumenical; see, for example (Carston, 2013).

elements. Thus, pragmatic processes are allowed to permeate what Minimalists standardly take to be *semantic* processes.

Like Minimalists and Indexicalists, most Contextualists reserve the descriptor ‘semantic’ for the processing of purely linguistic information, with all subsequent modulation and enrichment being treated as pragmatic.⁴⁹ However, they hold that semantic processing will typically generate only the ‘blueprint’ for a theoretically relevant meaning (Neale, 2007a). Thus, even though we can sensibly talk about sentences having minimal semantic meanings, these will typically play no role in our theorizing. Instead, Moderate Contextualists argue that the meanings of primary importance are the ones which speakers and hearers actually have in mind during communication. These ‘psychologically real’ meanings, which are reflected in our intuitive truth-value judgements, will typically depend on free pragmatic effects (although, on the odd occasion where no modulation or enrichment is required, they will coincide with minimal semantic meanings; however, this will be merely incidental – the exception rather than the rule).

The pragmatically enriched meanings posited by Contextualists have been conceptualised in various ways. In the next two sub-sections I will discuss Relevance Theory’s ‘explicatures’ (Sperber & Wilson, 1995) and Recanati’s notion of ‘what is said’ (Recanati, 2004, 2010). Both approaches classify a further set of meanings as pragmatic implicatures (thus Borg (2004) describes Contextualism as a form of ‘dual pragmatics’, positing two separate iterations of pragmatic effects). The question of precisely how to keep implicatures distinct from what is ‘said’ or ‘explicated’ will also briefly be considered.

6.1. Explicatures

Relevance Theory (RT) offers a psychological account of communication, according to which a modular process of linguistic decoding is distinguished from a pragmatic process of interpretation. Linguistic decoding is thought to generate an expression’s ‘logical form’ (LF), which includes the conventional meanings of its constituent expressions.⁵⁰ Crucially, Relevance Theorists claim that pragmatic processing operates on logical form in accordance with a

⁴⁹ The status of the composition process is rather unclear – is this supposed to be a semantic process that is interleaved with pragmatic ones, or is it seen as part of pragmatic processing? Although this is an interesting question, it lies beyond the scope of the current discussion.

⁵⁰ There is ongoing debate about the precise nature of these conventional meanings, including how information-rich they are. Carston proposes that they are extremely ‘thin’ (Carston, 2012, 2013, 2019). This represents a departure from the classic Fodorian conception of meanings as atomic concepts (Fodor, 1998). I will largely leave this issue aside, although it crops up again in the discussion of Radical Contextualism below.

principle of cognitive efficiency, known as the (communicative) principle of relevance.⁵¹ This principle states:

Every act of ostensive communication communicates a presumption of its own optimal relevance. (Sperber & Wilson, 1995, p. 158)

'Relevance' here is a technical term. To get a grasp on it we need to define two more basic notions. The first is the notion of an 'assumption', which can be thought of roughly as a belief: in Sperber and Wilson's words, assumptions are "thoughts treated by the individual as representations of the actual world (as opposed to fictions, desires, or representations of representations)" (Sperber & Wilson, 1995, p. 2). Second, 'contextual effects' are additions or changes to an individual's set of assumptions. With these concepts on the table, relevance is defined by the following 'extent conditions':

Extent condition 1: an assumption is relevant in a context to the extent that its contextual effects in this context are large.

Extent condition 2: an assumption is relevant in a context to the extent that the effort required to process it in this context is small.

(Sperber & Wilson, 1995, p. 125)

In less technical terms, an utterance's relevance depends (directly) on its impact on the hearer's beliefs and (inversely) on the effort she must expend in its interpretation.

The 'presumption of optimal relevance' that every utterance communicates is then defined by the following conditions (where intending to make an assumption manifest to someone can be thought of, roughly, as intending that they come to believe its content):

(a) The set of assumptions **I** which the communicator intends to make manifest to the addressee is relevant enough to make it worth the addressee's while to process the ostensive stimulus.

(b) The ostensive stimulus is the most relevant one the communicator could have used to communicate **I**.

(Sperber & Wilson, 1995, p. 158)

In other words, RT claims that utterances should be treated as though whatever the speaker intended to communicate is sufficiently informative to be worth the hearer's cognitive effort to

⁵¹ According to the separate *cognitive* principle of relevance, human cognition tends, in general, to be geared to the maximisation of relevance (Sperber & Wilson, 1995, pp. 260-261).

derive it. And, moreover, hearers can expect speakers to have packaged the information in such a way as to minimize the work required to retrieve it. This gives us a picture of relevance-based pragmatic processing as delivering the most readily available interpretation of an utterance that is compatible with its being informative enough to warrant the effort. We can think of hearers, then, as cycling through various possible interpretations of an utterance, beginning with those which are the most readily accessible and stopping as soon as they reach one that delivers cognitive effects worthy of the energy expended up to that point.

Relevance Theorists argue that both *explicatures* and *implicatures* are generated by relevance-based processing. *Explicatures* are conceptualised as ‘developments’ of the logical forms of uttered sentences, modulating and enriching these in accordance with the principle of relevance.⁵² *Implicatures*, on the other hand, are relatively unconstrained by logical forms; what is important is just that they can be inferred from the explicature, together with other background premises. To show how this is supposed to work, consider again the exchange considered earlier:

- A. There’s a lot of hype surrounding the match this afternoon.
- B. They won’t survive.

Given the scenario described above, in which A and B have been discussing the unseeded doubles team, A will derive *Kris and Marta* as the optimally relevant denotation of ‘they’ in B’s utterance. Similarly, A will derive the optimally relevant meaning of ‘survive’, as something like *progress through the tournament*. Given that A was talking about that afternoon’s match, the relevance principle will also result in the logical form being enriched with an unarticulated constituent, specifying that the pair won’t progress through *to the next round*. Thus, A derives the explicature in (1a), reproduced below.

(1a) Kris and Marta won’t progress through to the next round of the tournament.

Given an additional background assumption, along the lines that *supporters are disappointed when their team is knocked out*, (1a) justifies the implicature that *the supporters of the unseeded players will be disappointed*.

In Relevance Theory, both explicatures and implicatures are derived in parallel, with mutual adjustment between the two (Carston, 2002, pp. 142-152). In one direction, explicatures act as premises for implicatures. In the opposite direction, any implicature a hearer immediately hits

⁵² Bach (1994) identifies a similar class of meanings as ‘implicatures’ (note the third ‘i’: implic-i-tures, not implic-a-tures). These incorporate extra-linguistic effects aimed at delivering, first, *complete* propositions that are truth-evaluable and, second, *expanded* propositions that correspond to what is intuitively communicated.

on will require an explicature to be posited as a supporting premise. Thus, explicatures and implicatures are held to be importantly intertwined.

Although Relevance Theory allows both explicatures and implicatures to be affected by all kinds of extra-linguistic factors, operating freely of linguistic requirements, the two kinds of meaning are held apart by various criteria. As mentioned, explicatures are closely related to logical forms, being ‘developments’ or ‘expansions’ of these, whereas implicatures are not (Sperber & Wilson, 1995, p. 182). Explicatures have also traditionally been understood to survive certain logical embeddings that are supposed to block implicatures, including in the antecedents of conditionals. The associated diagnostic tool is known as the ‘scope test’ (Carston, 1998; 2002, pp. 191-194; Recanati, 1989). More recently, Hall (2008) and Carston and Hall (2012) have argued for a ‘locality constraint’ on explicatures, whereby modulation and enrichment must operate on sub-propositional constituents. In contrast, implicatures are held to be derived on the basis of complete propositional meanings. Carston (1998) also proposes a functional difference between explicatures and implicatures: whereas explicatures can act as premises from which implicatures are inferred, the reverse is not true – explicatures cannot be entailed by implicatures (Carston, 1988; 2002, pp. 189-190). The effectiveness of each of these tests has been disputed. For example, Borg (2016, 2017) argues that none of them manages to isolate the category of meaning that Relevance Theorists want to capture. She proposes instead that explicatures are pragmatic meanings for which speakers are held *conversationally liable*. Although I will not take a view on this debate, I will deploy each of the proposed tests in chapter 8 in considering whether the information conveyed by a frame could be classified as an explicature.

6.2. What is Said

Another prominent Contextualist account is that of Recanati (2010). According to his ‘Truth-Conditional Pragmatics’ (TCP), the truth-value judgements that underlie context shifting arguments are tracking what is ‘said’ by an utterance. He argues that what is ‘said’ is derived via ‘primary’ pragmatic processes, whereas implicatures are derived via ‘secondary’ pragmatic processes. Primary pragmatic processes are held to take place at a sub-personal, unconscious level. Crucially, they include the free effects of modulation and enrichment, in addition to the mandatory ‘saturation’ of context-specific elements in language. As Recanati puts it:

[Truth-Conditional Pragmatics] says that truth-conditional content may be affected not only by saturation (as when an indexical is assigned a contextual value) but also by *free* pragmatic processes (Recanati, 2010, p. 5, emphasis in original).

According to Recanati's 'availability principle', "what is said must be consciously available to the interpreter" (Recanati, 2004, p. 17). This prevents minimal semantic meanings from counting as what is 'said', except on the odd occasion where nothing more than mandatory saturation is required to retrieve what the speaker intuitively 'said'; in all other instances, hearers only have conscious access to some richer representation.

According to Truth Conditional Pragmatics, secondary pragmatic processes operate on what is 'said'. These processes are conceptualised as being consciously accessible and more richly inferential than primary processes.⁵³ Thus, although Recanati has appealed to some of the same criteria as Relevance Theory, discussed above, he ultimately grounds the distinction between what is 'said' and implicated in a difference of psychological processing. His view has also attracted criticism, though, from those who doubt whether it can ground a distinction between what is 'said' and what is implicated (Borg, 2016, 2017; Carston, 2002, pp. 166-170).

6.3. Radical Contextualism

Radical Contextualists make an in-principle objection to Minimalism: they claim that sentences themselves fail to express complete, truth-evaluable propositions.⁵⁴ Instead, it is argued, extra-linguistic factors always play a substantial and pervasive role. Thus, Travis writes:

The new view might be called the speaking-sensitive view of words, or language...[Words'] fixed, language-contributed semantics must, in general, be supplemented if they are to be properly assessable as to truth; that is, if they are to count either as true or as false. Nor is such supplementation provided automatically by some fixed stock of 'indexical features'. (Travis, 2008, p. 70)

In chapter 5 I will explore this view in detail and argue against the Radical Contextualist's in-principle rejection of minimal semantic meanings. I will conclude there that the more pertinent objection to Minimalism is the one made by Moderate Contextualists, that minimal semantic meanings are theoretically otiose. In chapter 8, however, I will argue against that claim too, on the basis that minimal semantic meanings are necessary for explaining framing puzzles.

⁵³ Carston (2009) interprets Recanati as claiming that reasoning about speaker intentions never comes into primary pragmatic processing, only secondary pragmatic processing. I think this interpretation must be mistaken, though, since Recanati takes even mandatory saturation to involve appeal to speaker intentions. For example, he writes: "the semantic value of context-sensitive expressions cannot be determined without appealing to fully-fledged pragmatic factors (e.g. speaker's intentions)" (Recanati, 2010, p. 2).

⁵⁴ Both Carston and Recanati appear to endorse this more radical stance in some of their writings: see for example, (Carston, 2002, 2013; Recanati, 2004, 2017).

6.4. Contextualism and Framing

From a Contextualist perspective, it might be proposed that alternative frames generate distinct explicatures (or, in Recanati's terminology, 'say' different things). In chapter 8 I will consider – and reject – this possibility. I will argue there that the distinction we need is one that separates a frame's minimal truth-conditional semantic meaning from the additional information it pragmatically implicates.

7. Semantic Relativism

Semantic Relativism (or, simply, 'Relativism') takes a different approach to context shifting arguments: rather than positing enriched meanings, it expands the set of parameters relative to which those meanings are evaluated for truth or falsity. This kind of approach has been advocated by Lasnik (2005), Predelli (2005), and MacFarlane (2009, 2014). While Relativists recognise a set of meaning entities similar to minimal semantic meanings, they deny that these have constant truth-conditions. To clarify the idea, I will first outline the standard view in formal semantics to which relativists are reacting.

Following Kaplan, formal semanticists have tended to conceptualise propositions as functions from a set of contextual parameters to a set of truth-conditional contents.⁵⁵ Truth-conditional contents, in turn, are understood as functions from possible worlds (or, perhaps, world-time pairs) to truth-values (i.e. *TRUE*, *FALSE*, and perhaps some third value). The set of parameters relative to which propositions are evaluated for truth are known as 'circumstances of evaluation' (Kaplan, 1989).

Relativists have sought to expand the circumstances of evaluation to include various non-standard parameters, which are freely brought into play by extra-linguistic, occasion-specific features. For example, consider again the following sentence, used of King John of England:

(2) John was old when he died.

The Relativist might analyse the proposition expressed by (2) as a function from circumstances of evaluation that include a comparison class parameter. If the value of this parameter is set to *English males alive in 2020*, (2) is mappable to *FALSE*. However, if the value is set to *English males alive in the middle ages*, (2) is mappable to *TRUE*. This is despite the fact that (2) itself is thought to have a constant meaning. According to the most extreme versions of Relativism, truth-conditions can be relativized to individual assessors, or point of assessment, not merely to

⁵⁵ As we have seen, the set of contextual parameters is conceptualised in different ways under different theories: for Minimalists and Indexicalists they include only what is required to saturate context-sensitive elements; for Contextualists, they may also include other occasion-specific extra-linguistic features.

certain non-standard parameters determined by the utterance context (Laserson, 2005; MacFarlane, 2014).

7.1. Relativism and Framing

Adopting a Relativist approach, it might be argued that the truth-conditions of alternative frames can differ, even while they share the same propositional meaning. In chapter 8 I will consider – and reject – this view. I will argue that, ultimately, it turns out to be either implausible or unparsimonious to appeal to such a parameter in explaining framing effects.

8. Occasion-Sensitivity

Both Contextualists and Relativists standardly assume that it is possible to represent a ‘sharp’ or ‘precise’ truth-conditional content, which captures what is intuitively expressed by a speaker’s utterance. Their opponents have sought to problematise this claim, pointing out that there is no principled way to establish how precise the representation ought to be. Instead, we could go on sharpening it indefinitely, in order to address ever more inventive context shifting arguments (Borg, 2012, pp. 35-38). Consider an utterance of (6):

(6) Mitch opened a book and began to read out loud.

In most contexts, (6) is intuitively true just if Mitch began to read out loud *from the book he opened*. Thus, the Contextualist may identify the following enriched content as the explicature of (6) (or what is ‘said’):

(6a) Mitch opened a book and began to read out loud from it.

However, now imagine that what Mitch actually did was select one word from each page and articulate it (perhaps starting with the final letter and proceeding backwards to the first letter, so that ‘about’ would be pronounced ‘t-u-o-b-a’). The sharpened formulation in (6a) may be judged true or false of this state of affairs, depending on the context. In many contexts, (6a) would be judged false, given that Mitch’s behaviour is not what we would normally think of as *reading*. In some contexts, though, it might be judged true – say, if it is a response to the question ‘What caused the disruption in the library today?’. This points to a need to further sharpen the content in order to capture exactly what Mitch did. It is unclear, though, how the Contextualist can ever terminate the sharpening process, in order to specify once and for all the communicated content.

This kind of worry is avoided by the more radical position of Occasion-sensitivity advocated by Travis (2000, 2006, 2008). According to Occasion-sensitivity, extra-linguistic factors play an *ineliminable* role in meaning-determination. It is held to be impossible, even in principle, to

represent what is said linguistically (or in any other representational format). Instead, meanings are always constituted in part by the wider situation. Specifically, Travis understands them to be essentially related to what speakers are trying to *do*. The theory thus draws heavily on the Wittgensteinian conception of *meaning as use*. Travis endorses the following claim, which he ascribes to Wittgenstein:

Content is inseparable from point. What is communicated in *our* words lies, inseparably, in what we would expect of them. How our words represent things is a matter of, and not detachable from, their (recognizable) import for our lives (Travis, 2006, p. 33, emphasis in original).

In this way, he rejects altogether the standard conception of utterances, as casting propositional ‘shadows’ that could be represented in language – or, at least, in thought (Travis, 2000).

Ultimately, the question of whether it is possible to accurately represent the communicated meaning of an uttered sentence is one which falls outside the scope of the thesis (as will be discussed again towards the end of chapter 5). I will therefore end up focusing, in chapter 8, on a less radical claim that is shared between Contextualism and Occasion-Sensitivity – namely that meanings with theoretical import will generally be context-dependent (whether or not they could, in principle, be captured representationally).

9. Conclusion

The purpose of this chapter has been to survey the spectrum of philosophical approaches to contextual variations in meaning. I will now begin to apply these approaches to the case study of framing. I begin in the next chapter by considering whether framing effects can be traced back to a basic semantic asymmetry.

CHAPTER 3

Inexhaustive Alternatives

Abstract

In this chapter, I examine a deflationary approach to framing effects, which attempts to resolve framing puzzles by *dissolving* them. The hypothesis is that alternative frames provoke distinct representations in hearers because of their basic semantic inequivalence, traceable back to the conventional meanings of their constituent expressions. I observe that the supposedly contradictory predicate expressions substituted across pairs of framing conditions in the classic ‘Asian Disease Problem’ – ‘be saved’ and ‘die’ – are not jointly exhaustive. I report on two new experiments that investigate whether this fact could be driving the framing effect. The results suggest that the framing effect remains robust even with jointly exhaustive predicate expressions. Moreover, framing effects have been shown to arise in various other studies that use genuinely contradictory predicate expressions. I conclude by rejecting the deflationary hypothesis.

1. Introduction

In chapter 1, I introduced a classic risky-choice framing paradigm known as the ‘Asian Disease Problem’ (ADP), which was first presented in (Tversky & Kahneman, 1981). This is reproduced below:

Imagine the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume the exact scientific estimates of the consequences are as follows:

Framing condition 1:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.

Framing condition 2:

If Program C is adopted, 400 people will die.

If Program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.

I will now examine more closely the properties of the sentences used in the two framing conditions. These are supposed to comprise two pairs of alternative frames, which describe the same prospects. I reorganise the four sentences below, to pair up (1) with (2), and (3) with (4):

(1) If Program A is adopted, 200 people will be saved.

(2) If Program C is adopted, 400 people will die.

(3) If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.

(4) If Program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.

The equivalence of the alternative frames depends on two factors working in tandem: first, the meanings of the predicate expressions 'be saved' and 'die' are supposed to be *contradictory*, denoting sets that are *mutually exclusive* and *jointly exhaustive*. Second, the number expressions in corresponding clauses are supposed to be *complementary*, denoting numbers that *sum to 600*.⁵⁶ Together, these factors are supposed to ensure that the meanings expressed by the alternative frames are *logically equivalent* to each other. In other words, they are supposed to share a single set of truth-conditions, such that (1) will be true just in case (2) is; and (3) will be true just in case (4) is.

In this chapter, I will challenge the assumption that the predicate expressions have contradictory conventional meanings. Specifically, 'be saved' and 'die' denote sets of entities that are not jointly exhaustive.⁵⁷ In the next chapter, I will challenge the assumption that the number expressions are complementary. For now, though, I will simply assume that they are.

The structure of the chapter is as follows: in §2, I briefly survey various respects in which the alternative frames in the ADP fail to be logically equivalent. In §3 I focus on the fact that 'be saved' and 'die' are not jointly exhaustive, and I explain how this fact could potentially be driving the classic framing effect. In §4 I discuss existing experimental findings, which offer some *prima facie* support for such a deflationary hypothesis. However, in §§5-6 I present new data that suggests the hypothesis is false. In §7, I provide examples of frames that are not

⁵⁶ So, for example, with respect to (1) and (2) $200 + 400 = 600$. Meanwhile, in the first clauses of (3) and (4), $600 + 0 = 600$; and in the second clauses, $0 + 600 = 600$.

⁵⁷ This deflationary challenge is incompatible with Occasion-sensitivity and radical forms of Contextualism and Relativism, which deny that expressions can conventionally denote sets of entities at all. I will confront that issue in chapter 5. The hypothesis discussed in the current chapter remains compatible with Minimalism, Indexicalism, and moderate forms of Contextualism and Relativism.

subject to the deflationary challenge. Finally, in §8, I argue that framing effects cannot be analysed as ‘Frege puzzles’.

2. Clearing the Way

Before getting to the heart of the deflationary challenge, I will briefly discuss a series of related concerns about the frames used in the ADP. It is fairly clear that, out of context, the meanings of (1) and (2) are not logically equivalent.⁵⁸ First, the number expressions can only be considered complementary once the domain is restricted to the 600 people who are expected to be killed by the disease. This is not guaranteed by the frames themselves, taken out of context. It is possible, then, that when these sentences are presented, hearers represent the domain as being larger or smaller than 600. Indeed, Okder (2012) proposes that some participants fail to think of ‘200’ and ‘400’ in (1) and (2) as *subsets of the 600 people* referred to in the vignette. Although this is an interesting hypothesis, I will not consider it further here.⁵⁹ I will assume instead that (1) and (2) are straightforwardly elliptical, with the vignette supplying the additional information shown in italics below:

(1a) If Program A is adopted, 200 people *out of 600 expected to be killed by the outbreak of an unusual Asian disease* will be saved.

(2a) If Program C is adopted, 400 people *out of 600 expected to be killed by the outbreak of an unusual Asian disease* will die.

I assume, then, that the predicate expression ‘be saved’ denotes a *subset* of the 600 people expected to be killed by the disease – namely those who are saved. And I assume that ‘die’ denotes the subset who die.

A further issue, though, is that the two subsets are not *mutually exclusive*. After all, given that *all* people die eventually, anyone who is saved is also someone who dies. In other words, the set of individuals who are saved is a subset of those who die. This prevents the conventional meanings of ‘be saved’ and ‘die’ from being straightforwardly contradictory. That said, I take it to be sufficiently clear from the vignette that what is meant is expressible as in (1b) and (2b):

(1b) If Program A is adopted, 200 people out of 600 expected to be killed by the outbreak of an unusual Asian disease will be saved *from the disease*.

⁵⁸ Similar points apply to (3) and (4). However, I will focus on (1) and (2) since they are linguistically simpler.

⁵⁹ I believe there are methodological and interpretational problems with Okder’s experiments. Moreover, the hypothesis is inapplicable to the studies I discuss in §7 below. Therefore, even if it helps explain the framing effect in the ADP, it cannot account for framing effects in general.

(2b) If Program C is adopted, 400 people out of 600 expected to be killed by the outbreak of an unusual Asian disease will die *from the disease*.

I will proceed as though the experimenters had presented their participants with the linguistic stimuli in (1b) and (2b). However, I will now suggest that even these more specific formulations remain logically inequivalent.

3. The Deflationary Hypothesis

The observation I want to pursue is that the sets denoted by 'be saved' and 'die' need not jointly exhaust the domain. Specifically, some of the 600 people expected to be killed by the disease could end up staying alive without being saved by the relevant program. That might be, for example, because they do not contract the disease in the first place, or because they unexpectedly recover from it, without the assistance of the proposed intervention. In any possible world that contains such individuals, (1b) and (2b) will have different truth-values: (1b) would express something true (since it is consistent with 200 people being saved that some or all of the remaining 400 stay alive), whereas (2b) would express something false (since it is inconsistent with 400 people dying that any of these 400 stay alive). This shows that the truth-conditions of the two sentences must diverge: they must have different intensions on account of having different extensions in some possible worlds. In sum, since more people could live under Program A than Program C, the sure option in the 'be saved' condition is strictly better than the sure option in the 'die' condition.⁶⁰

The same reasoning applies to the risky options in (3) and (4).⁶¹ Specifically, the second complements of those options, in the final clauses of the sentences, are inequivalent. Under (3), there is a two-thirds probability that none of the 600 people will be saved; however, it remains possible that some (or all) may nevertheless stay alive, for the reasons set out above. Under the second complement of (4), however, this is impossible, since there is a two-thirds probability that all 600 will *die*. This means that (the second complement of) the risky option in the 'be saved' condition is strictly better than that in the 'die' condition. In sum, both the sure and risky options in the 'be saved' condition are strictly better than their counterparts in the 'die' condition.

⁶⁰ A similar point is made by Berkeley and Humphreys (1982). However, they focus on the possibility that more than 200 end up being *saved*, rather than the possibility that some of the remaining 400 might live *without* needing to be saved. Their critique is therefore more closely allied to the proposal I discuss in the next chapter.

⁶¹ Note that I am assuming 'no people' in the 'be saved' condition to be straightforwardly co-extensional with 'nobody' in the 'die' condition.

Note that the wider discourse context does not obviously secure the joint exhaustivity of 'be saved' and 'die'. No doubt it was the experimenters' intention to constrain the problem in this way. However, it does not follow from the vignette that people will certainly die if they are not saved; otherwise the following argument should hold:

Premise 1: A disease is expected to kill 600 people.

Premise 2: Two alternative programs to combat the disease have been proposed.

Conclusion: Any of the 600 people who are not saved by one of the two programs will die.

On the contrary, this argument is clearly invalid. It fails on two separate counts: first, given that expectations need not be met, it is not possible to conclude from the first premise that *any* number of people will die in the absence of intervention. There is nothing in the vignette which instructs participants to treat the expected death toll as a certainty.⁶²

Relatedly, the second premise does not require that the two alternative programs are the *only* possible factors affecting patterns of survival. Instead, other deliberate interventions, or accidental events, could change the expected death toll.

This basic semantic asymmetry between the conditions of the ADP opens up the possibility that the apparent framing effect is being driven purely by the fact that more lives might be saved in the 'be saved' condition than in the 'die' condition. The wider literature provides some grounds for predicting this. First, a meta-analysis of risky choice framing effects, conducted by Kühberger et al. (1999), finds that risk aversion is positively correlated with the *size* of the payoff, especially in the gain domain. In other words, as the number of lives to be saved increases, people tend to become more risk-averse.

Perhaps, then, the standard pattern of responses to the ADP might be glossed as follows: regardless of framing condition, participants represent the options as potential *gains* (relative to 600 people being killed by the disease). However, the potential gain is *larger* in the 'be saved' condition than in the 'die' condition, since *more* people could stay alive. Due to the potential gain being larger in the 'be saved' condition, participants are more risk-averse in that condition, explaining their greater preference for the sure option. Conversely, in the 'die' condition, where

⁶² Bohm and Lind (1992) make a similar point. They test a manipulation of the ADP which recasts the options in terms of *expectations*, in line with the vignette, rather than *certainties* (as in the original conditions of the ADP). For example, the sure option in the 'be saved' frame is rephrased as follows: 'If Program A is adopted, *it is expected that* 20 people will be saved.' (Note that the experimenters relocated the ADP scenario from the US to Sweden and reduced the number of lives by a factor of ten, to reflect the smaller population there.) However, framing effects are found to persist even under this variation of the ADP.

the potential gain is smaller, they are less risk-averse, and tend to shift towards the risky option. This hypothesis could potentially explain the standard pattern of ‘shifty’ responses to the ADP. It therefore serves as a competitor to the prospect theoretic account, which assumes that outcomes are represented as gains in the ‘be saved’ condition and losses in the ‘die’ condition. Rather than the shift in preferences being a genuine framing effect, it could be explained purely in terms of the magnitude of the potential gain.

An alternative hypothesis that makes the same prediction is that recipients of the ADP tend to focus on the sure option rather than the risky one, as Kühberger and Gradl (2013) suggest. These researchers replicate earlier findings that framing affects how participants rate the attractiveness of the sure option but not the risky option. In a series of further experiments, they present evidence that framing effects are driven mainly by the contrast between the sure options in each condition.⁶³ Since, as I have argued, the sure option in the ‘be saved’ condition outperforms the sure option in the ‘die’ condition, this could explain why it is evaluated more positively in the ‘be saved’ condition than in the ‘die’ condition.

It is also worth noting that there is now a categorical difference between the outcomes afforded by each framing condition: in the ‘be saved’ condition, both the sure and risky options are strictly consistent with all 600 people staying alive. In contrast, in the ‘die’ condition, only the risky option is consistent with that outcome. Thus, if participants were motivated to leave this possibility open, they would be forced to select the *risky* option in the loss condition. That could be another reason to expect that the sure option will be more popular in the ‘be saved’ condition than in the ‘die’ condition.

Each of these three hypotheses might explain the standard pattern of responses generated by the ADP, without making any appeal to the (positive or negative) *valence* of the wording in each framing condition. In this sense they would *dissolve* the framing effect. That is not to say that there would be no interesting behaviour left to explain, of course, since we would still need to build the effects of payoff size and so on into our model of decision-making.⁶⁴ The point is just that the explanation would not appeal to *wording* choices, which characterise framing effects.

⁶³ This effect may be exacerbated by the sure option typically being presented *before* the risky option. In other words, there could be a primacy effect, whereby participants focus on whichever option is presented first. However, there is mixed evidence for a primacy effect: while Kühberger & Gradl find an order effect in their second experiment, their third experiment suggests that this is not the main driver of the framing effect.

⁶⁴ It would also be necessary to explain why frames tend to be represented as equivalent when both are made available. I do not consider this issue further here, since I will ultimately reject the deflationary strategy altogether.

Interestingly, this particular challenge does not appear to have been explored in the framing literature. Instead, a plethora of studies following in the wake of Tversky and Kahneman's continue to contrast 'be saved' with 'die': these include, for example, (Chick, Reyna, & Corbin, 2016; Frisch, 1993; Jou, Shanteau, & Harris, 1996; Kühberger, 1995; Mandel, 2001, 2014; Van Schie & Van Der Pligt, 1995). This leaves a large portion of the framing literature potentially vulnerable to the deflationary challenge. It is therefore important to establish whether that challenge could succeed. The first step is to test whether the 'shifty' responses to the ADP really do depend on the semantic asymmetry between the two framing conditions. If they do, the framing effect ought to be eliminable by using predicate expressions with genuinely contradictory conventional meanings.

4. Negation

Perhaps the most obvious way of ensuring that pairs of predicate expressions have jointly exhaustive denotations would be to use straightforward negation: for example, 'be saved' could be contrasted with 'not be saved'; or 'die' with 'not die'. Mandel (2001) manages to eliminate framing effects using exactly this strategy.

Mandel tests various re-wordings of the ADP options, four of which are relevant to the current discussion. The first presents the options purely in terms of the numbers of people who will 'be saved'. The second describes the numbers who will 'not be saved'. The third presents the options in terms of the numbers of people who will 'not die'. The fourth describes the numbers who will 'die'. The re-worded conditions are set out below:

- (1) If Plan A is adopted, it is certain that 200 people will be saved.
If Plan B is adopted, there is a one-third probability that all 600 people will be saved.
- (2) If Plan A is adopted, it is certain that 400 people will not be saved.
If Plan B is adopted, there is a two-thirds probability that all 600 people will not be saved.
- (3) If Plan A is adopted, it is certain that 200 people will not die.
If Plan B is adopted, there is a one-third probability that all 600 people will not die.
- (4) If Plan A is adopted, it is certain that 400 people will die.
If Plan B is adopted, there is a two-thirds probability that all 600 people will die.

There are two important points to note about Mandel's reformulations: first, each includes only one or other complement of the risky option, whereas the classic ADP conditions present both.

This is an unavoidable consequence of limiting the linguistic construction to be either affirmative or negative, rather than being a mixture of each. Second, Mandel adds the phrase ‘it is certain that’ into each sure option. This is intended to discourage participants from thinking that the actual numbers being saved or dying could be different from those specified.

Crucially, Mandel finds no significant differences between participants’ responses in conditions (1)-(4). In other words, framing effects were eliminated by straightforwardly negating the predicate expressions. The data are therefore consistent with the deflationary hypothesis: the apparent framing effect may depend on the asymmetry in the classic ADP between conditions that contrast ‘be saved’ with ‘die’.

There are, however, several obstacles to treating Mandel’s data as supporting the deflationary hypothesis. First, his inclusion of the phrase ‘it is certain that’ in the sure options introduces a new confound; perhaps this is what eliminates the framing effect (for example by reducing the possibility that *more* people could be saved – I will discuss this possibility in detail in chapter 4). Second, the curtailment of the risky options could also be doing some work, by failing to constrain the alternative outcomes. Third, negated constructions are morphologically more complex, and therefore potentially harder to process, than affirmative ones. This may prevent Mandel’s conditions (2) and (3) from being directly comparable to (1) and (4). In sum, it is difficult to provide a robust explanation of why framing effects were eliminated in this experiment.⁶⁵

It is also notable that Mandel found no evidence of a framing effect arising from the use of ‘be saved’ (in conditions (1) and (2)) as opposed to ‘die’ (in conditions (3) and (4)). This looks like a troubling result for the proponent of the deflationary strategy, since the original effect is hypothesised to depend on those terms failing to be jointly exhaustive.

Clearly, Mandel’s data cannot be straightforwardly marshalled in support of the deflationary challenge. A better way to test the hypothesis directly would be to replace the pair of predicates used in the classic ADP with ones that are more plausibly contradictory. In order to plug a gap in the framing literature, this strategy is pursued in two new experiments, the details of which are reported below. The experiments were conducted in collaboration with Nat Hansen and Kathryn B. Francis. Since the findings have not been published elsewhere, I will provide a full description of the experimental method and results in the next two sections.⁶⁶ Before getting

⁶⁵ This is not a criticism of Mandel’s study, since he conducted it with different objectives in mind: specifically, he was interested in establishing the extent to which framing effects depend on the use of affirmative or negated constructions, as opposed to positively- or negatively-valenced predicate expressions.

⁶⁶ The data are also openly available from the University of Reading Research Data Archive, at <http://dx.doi.org/10.17864/1947.245>.

into the details, it is worth emphasising the philosophical motivation of the research upfront. The experiments are designed to investigate whether a basic semantic asymmetry between the framing conditions, resulting in a potentially important logical distinction, can explain why people respond differently under each condition.

5. Experiment 1

In the first experiment, which acted as a pilot study, two additional conditions were tested alongside the two classic conditions of the ADP. The new conditions used the expressions ‘live’ and ‘survive’ in place of ‘be saved’. When contrasted with ‘die’, these expressions more plausibly denote exhaustive sets.⁶⁷ The reformulated options are thus designed to bring the two conditions of the ADP closer to equivalence. For example, if only 200 out of 600 people will live, or survive, this now implies that the remaining 400 will certainly die. To the extent that the framing effect depends on the semantic mismatch between ‘be saved’ and ‘die’, this manipulation ought to undermine it. Therefore, the hypothesis to be tested is that the standard cross-condition shift in preferred option will be attenuated or eliminated where the first framing condition uses ‘live’ or ‘survive’ in place of ‘be saved’.

5.1. Method

5.1.1. Participants

An a priori power analysis indicated that a total of 273 participants would be required in order to detect a medium to large framing effect (of Cramer’s $V = 0.2$).⁶⁸ To account for screening processes, an additional 10-15% were recruited. As such, a total of 305 participants were recruited through Amazon Mechanical Turk and paid \$0.40 for their participation. However, 63 responses (comprising 21% of the data) were screened out for one or more of the following reasons:

- Participants said they had encountered the ADP before (35 responses).
- They failed the screener question (14 responses).⁶⁹

⁶⁷ This is perhaps less clear with ‘survive’ than with ‘live’, since it is possible that some people never contract the disease in the first place, and therefore can’t be said to ‘survive’ it. I return to this point later.

⁶⁸ In line with Cohen (1988, p. 222), when $df=3$, effect sizes are interpreted as follows: an effect of Cramer’s $V=0.06$ counts as small; an effect of Cramer’s $V=0.17$ counts as medium; and an effect of Cramer’s $V=0.29$ counts as large. We expected the ADP to generate a relatively large overall framing effect, based on the meta-analysis by Kühberger (1998). However, we expected the effects to be a little smaller for the pairwise comparisons between the individual conditions. Since there were no clear precedents for this in the literature, we decided in the first instance to target a medium-large effect size of Cramer’s $V=0.2$.

⁶⁹ After completing the main task, all participants received the following screener question, which checked for attention and basic numerical ability:

- They said they did not have English as a first language (7 responses).
- They did not complete the survey (4 responses: 2 dropped out prior to consent; 2 afterwards).
- They completed the survey too fast (1 response, which was completed in 17 seconds).⁷⁰
- They were responding from a VPS, non-US, or unknown IP type and their responses to the open-ended question was irrelevant or nonsensical, or the responses were probable duplicates (10 further responses) – see Appendix A for further details of this screening procedure.

This left 242 participants.⁷¹ The numbers in each experimental condition were as follows:

- ‘be saved’ condition: 55 participants.
- ‘survive’ condition: 62 participants.
- ‘live’ condition: 62 participants.
- ‘die’ condition: 63 participants.

The final dataset therefore ended up being slightly under-powered, due to more responses needing to be screened out than expected.

5.1.2. Design

There were four between-subjects framing conditions (‘be saved’, ‘survive’, ‘live’, and ‘die’) and an independent variable of option (sure vs. risky).

5.1.3. Procedure and materials.

Participants were randomly assigned to one of the four framing conditions, in roughly equal proportions. The only variation from the original problem was the addition of two new versions of the first condition, using ‘survive’ or ‘live’ instead of ‘be saved’, as follows:

If Program A is adopted, 200 people will (survive) (live).

Now imagine you are considering two jobs.
 If you take job A, you will be paid \$9 per hour.
 If you take job B, you will be paid \$15 per hour.
 Which job pays more per hour? [Job A/Job B]

⁷⁰ The minimum response time for this experiment was set at 20 seconds, based on a reading speed of 250-300 words per minute. This figure was derived from Human Interaction Speeds published at the following link (accessed in October 2017): <https://www.brainkart.com/article/Human-Interaction-Speeds-9017/>.

⁷¹ 54.5% female, $M_{\text{age}} = 36.5$ years, $SD_{\text{age}} = 11.9$ years.

If Program B is adopted, there is a one-third probability that 600 people will (survive) (live), and a two-thirds probability that no people will (survive) (live).

The experiment was run online, using Qualtrics/ Amazon Mechanical Turk. Participants were required to complete each question before they could advance. They could not go back to review or change answers on previous pages.

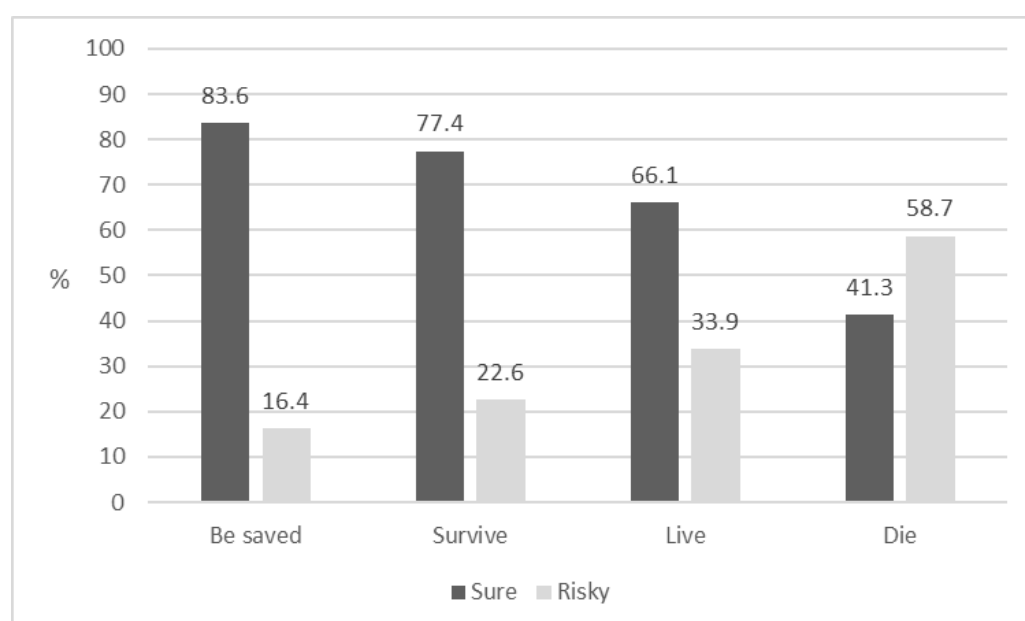
5.1.4. Analysis

Statistical analyses were performed in SPSS. A Chi-square test was performed in order to examine the overall relationship between framing condition and option selection. Cramer's V was also calculated to establish the size of this effect. Adjusted residuals were calculated for each option in each condition, in order to see what was driving the overall association. Six pairwise comparisons were run, using two-way Chi-square tests with Bonferroni corrections, to establish which conditions were significantly different from one another. Odds ratios were calculated manually to compare the sizes of the differences.

It was hypothesised that there would be a significant overall association between framing condition and option, driven mainly by the classic ADP framing conditions. It was further hypothesised that the framing effect would be larger with respect to the classic ('be saved') version of the first framing condition relative to the new ('survive' / 'live') versions, such that there would be a greater tendency to select the sure option in the classic version than in these other two versions.

5.2. Results

The results are summarised in the chart below:



A Chi-square test was performed to determine whether there was an association between framing condition and option (sure vs. risky). The analysis revealed a significant association ($\chi^2=28.59$), $p<.005$, $df=3$). The effect was large (Cramer's $V=.34$).

The full contingency table is provided below:

| | | | Option | |
|-------------------|----------|-------------------|--------|-------|
| | | | Sure | Risky |
| Framing condition | Be Saved | Count | 46 | 9 |
| | | Expected Count | 36.6 | 18.4 |
| | | Adjusted Residual | 3.1* | -3.1* |
| | Survive | Count | 48 | 14 |
| | | Expected Count | 41.2 | 20.8 |
| | | Adjusted Residual | 2.1* | -2.1* |
| | Live | Count | 41 | 21 |
| | | Expected Count | 41.2 | 20.8 |
| | | Adjusted Residual | -.1 | .1 |
| | Die | Count | 26 | 37 |
| | | Expected Count | 41.9 | 21.1 |
| | | Adjusted Residual | -4.9* | 4.9* |
| Total | | | 161 | 81 |

*indicates counts that are significantly different from expected counts ($z>1.96$).

In order to assess which conditions were driving the overall association between framing condition and option, 'adjusted residuals' were calculated. These tell us whether the proportion of participants choosing each option in each framing condition differed significantly from the 'expected' proportion. Expected proportions are, in turn, generated by assuming that each condition follows the overall distribution (in which 66.5% of respondents favoured the sure option, and 33.5% favoured the risky option).

As shown in the table above, the adjusted residuals were significant in three conditions: the 'be saved', 'survive', and 'die' conditions. In the 'be saved' and 'survive' conditions, significantly more people than expected chose the sure option ($z=3.1$ in the 'be saved' condition; $z=2.1$ in the 'survive' condition). By the same token, significantly fewer people than expected chose the risky option in those conditions. In the 'die' condition, significantly more people than expected chose the risky option ($z=4.9$), with significantly fewer than expected choosing the sure option. In the 'live' condition the differences were not significant ($z<1.96$). In summary, the association between framing condition and option was mainly driven by three of the four conditions: the 'be saved', 'survive', and 'die' conditions.

Follow-up analysis was performed to directly compare pairs of conditions. Six pairwise comparisons were conducted using two-way Chi-square tests with Bonferroni corrections. The results are shown in the table below:

| Comparison | χ^2 | Cramer's V | p-value |
|-------------------------|----------|------------|---------|
| 'Be saved' vs 'die' | 22.16 | .43 | <.008 |
| 'Survive' vs 'die' | 16.91 | .37 | <.008 |
| 'Live' vs 'die' | 7.76 | .25 | <.008 |
| 'Be saved' vs 'Live' | 4.69 | .20 | .04 |
| 'Be saved' vs 'Survive' | .71 | .08 | .49 |
| 'Survive' vs 'Live' | 1.95 | .13 | .23 |

The first three pairwise comparisons were significant at the Bonferroni-adjusted level of $p=.008$. This suggests that, in each version of the first framing condition (using 'be saved', 'survive' and 'live'), the pattern of responses differed significantly from those observed in the second condition (using 'die'). In other words, framing effects were observed regardless of which version of the first framing condition was used. However, the size of the framing effects differed between the three versions. This is shown by the odds ratios in the first three rows of the table below. The largest effect was observed for the classic ('be saved') gain condition, followed by the 'survive' condition and then the 'live' condition.

| Comparison | Odds ratio (<i>i.e. how much higher were the odds of participants choosing the sure option</i>) |
|-------------------------|---|
| 'Be saved' vs 'die' | 7.3 |
| 'Survive' vs 'die' | 4.9 |
| 'Live' vs 'die' | 2.8 |
| 'Be saved' vs 'Live' | 2.6 |
| 'Be saved' vs 'Survive' | 1.5 |
| 'Survive' vs 'Live' | 1.8 |

The remaining three pairwise comparisons, among the 'be saved', 'survive', and 'live' conditions themselves, showed no significant differences at the Bonferroni-adjusted level. However, the odds of participants choosing the sure option were highest in the 'be saved' condition, followed by the 'survive' condition, then the 'live' condition, as shown in the last three rows of the table above. Most notably, participants were more than twice as likely to choose the sure option in the 'be saved' condition than in the 'live' condition. These results provide some evidence that the framing effect is attenuated – although not entirely eliminated – when 'be saved' is replaced with 'live' or (to a lesser extent) with 'survive'.

5.3. Discussion

The experiment found an overall framing effect: in the 'die' condition, preferences shifted towards the risky option, as compared with the other three conditions. However, the effect was smaller with respect to the new versions of the first framing condition, which used 'survive' and 'live' rather than 'be saved'. This suggests that the modified wording may have attenuated the

framing effect. As such, the results offer some preliminary support for the deflationary hypothesis, according to which framing effects are driven by the fact that 'be saved' and 'die' fail to jointly exhaust the logical space.

There is also some evidence to suggest that the framing effect is attenuated to a *greater* extent under the 'live' condition than under the 'survive' condition. This can also be explained consistently with the deflationary hypothesis. Describing people as 'surviving' suggests that they do, in fact, contract the disease; there is something they need to 'survive' rather than their simply continuing to stay alive. Therefore, even if only 200 'survive', some of the remaining 400 people might be thought to stay alive, if they never contract the disease in the first place – if there is nothing they have to 'survive'. In contrast, if only 200 of the 600 people 'live' then it is impossible for any of the remaining 400 to stay alive.

In summary, the results of the first experiment provide some preliminary evidence that the ADP framing effect arises, at least partly, due to the inexhaustive conventional meanings of 'be saved' and 'die'. However, the data in this experiment did not show significant differences in direct comparisons between the 'be saved' condition and the 'survive' and 'live' conditions. This may have been due to the relatively large number of pairwise comparisons performed, which required the level of significance to be reduced to $p=.008$. Therefore, a second experiment was run, which removed the intermediate 'survive' condition and retained the other three conditions (resulting in the number of pairwise comparisons being reduced from six to three). The purpose of this experiment was to establish whether there was a genuine difference between the 'be saved' and 'live' conditions.

6. Experiment 2

6.1. Method

6.1.1. Participants

An a priori power analysis indicated that a total sample size of 241 would be required to detect a medium sized framing effect (of Cramer's $V = 0.2$).⁷² To account for screening processes, an additional 25-30% were recruited. As such, 308 participants were recruited through Amazon Mechanical Turk and paid \$0.40 for their participation. However, 112 responses (comprising 36% of the data) were screened out for one or more of the following reasons:

⁷² In line with Cohen (1988, p. 222), when $df=2$, effect sizes are interpreted as follows: an effect of Cramer's $V=0.07$ counts as small; an effect of Cramer's $V=0.21$ counts as medium; and an effect of Cramer's $V=0.35$ counts as large. Based on the pilot study, we were targeting a medium effect size of Cramer's $V=0.2$.

- Participants said they had encountered the ADP before (69 responses).
- They failed the screener question (31 responses).⁷³
- They said they did not have English as a first language (2 responses).
- They were responding from a VPS, non-US, or unknown IP type and their responses to the open-ended question were irrelevant or nonsensical (30 further responses) – see Appendix A for further details of the screening procedure.

This left 196 participants.⁷⁴ The numbers in each experimental condition were as follows:

- 'be saved' condition: 60 participants.
- 'live' condition: 70 participants.
- 'die' condition: 66 participants.

Again, then, the final dataset ended up being slightly under-powered, due to more responses needing to be screened out than expected.

6.1.2. Design

There were three between-subjects framing conditions ('be saved', 'live', and 'die') and an independent variable of option (sure vs risky).

6.1.3. Procedure and materials

Participants were randomly assigned to one of the three framing conditions, in roughly equal proportions. The only variation from the original ADP was the addition of a version of the first condition which used 'live' instead of 'be saved' in the options, as follows:

If Program A is adopted, 200 people will live.

If Program B is adopted, there is a one-third probability that 600 people will live, and a two-thirds probability that no people will live.

The experiment was run online, using Qualtrics/ Amazon Mechanical Turk. Again, participants had to complete each question before they could advance, and they could not go back to review or change answers on previous pages.

⁷³ After completing the main task all participants were asked the same screener question as in the previous experiment.

⁷⁴ 53.1% male, $M_{\text{age}} = 38.6$ years, $SD_{\text{age}} = 11.8$ years.

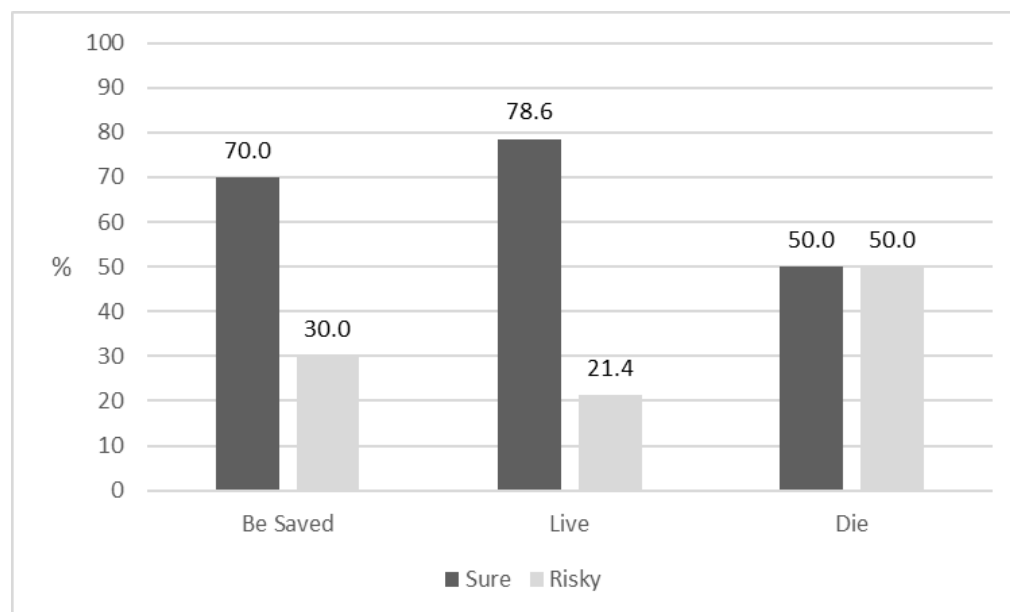
6.1.4. Analysis

As before, statistical analyses were performed in SPSS. A Chi-square test was performed to examine the overall relationship between framing condition and option selection. Cramer's V was calculated to establish the size of this effect. Adjusted residuals were calculated for each option in each condition, in order to see what was driving the overall association. Three pairwise comparisons were run, using two-way Chi-square tests with Bonferroni corrections, to establish which conditions were significantly different from one another. Odds ratios were calculated manually to compare the sizes of the differences.

It was hypothesised that there would be a significant overall association between framing condition and option, driven mainly by the classic ADP framing conditions. It was further hypothesised that the framing effect would be larger with respect to the classic ('be saved') version of the first framing condition than the new ('live') version, such that there would be a greater tendency to select the 'sure' option in the classic version than in the new version.

6.2. Results

The results are summarised in the chart below:



The Chi-square test was performed to determine whether there was an association between framing condition and (sure vs. risky) option. The analysis revealed a significant association ($\chi^2=12.94$, $p<.005$, $df=2$). The effect was medium-sized (Cramer's $V=.26$).

The full contingency table is provided below:

| | | | Option | |
|-------------------|----------|-------------------|--------|-------|
| | | | Sure | Risky |
| Framing condition | Be Saved | Count | 42 | 18 |
| | | Expected Count | 39.8 | 20.2 |
| | | Adjusted Residual | .7 | -.7 |
| | Live | Count | 55* | 15* |
| | | Expected Count | 46.4 | 23.6 |
| | | Adjusted Residual | 2.7 | -2.7 |
| | Die | Count | 33* | 33* |
| | | Expected Count | 43.8 | 22.2 |
| | | Adjusted Residual | -3.4 | 3.4 |
| Total | | | 130 | 66 |

*indicates counts that are significantly different from expected counts ($z > 1.96$).

As in the previous experiment, adjusted residuals were calculated to assess which conditions were driving the overall effect.⁷⁵ As shown in the table above, the adjusted residuals were significant in the 'live' and 'die' conditions. In the 'live' condition, significantly more people than expected chose the sure option ($z = 2.7$). By the same token, significantly fewer people than expected chose the risky option ($z = -2.7$). In the 'die' condition, significantly more people than expected chose the risky option ($z = 3.4$), with significantly fewer than expected choosing the sure option. In the 'be saved' condition the differences were not significant ($z < 1.96$). In summary, in the second experiment the association between framing condition and option was mainly driven by the 'live' and 'die' conditions. This finding conflicts with that of the first experiment, and with the deflationary hypothesis.

As before, pairwise comparisons were conducted using two-way Chi-square tests with Bonferroni corrections. The results are shown in the table below:

| Comparison | χ^2 | Cramer's V | p-value |
|----------------------|----------|------------|--------------|
| 'Be saved' vs 'die' | 5.22 | .20 | .03 |
| 'Live' vs 'die' | 12.14 | .30 | <.017 (.001) |
| 'Be saved' vs 'Live' | 1.25 | .10 | .31 |

Only the second pairwise comparison was significant at the Bonferroni-adjusted level of $p = .017$. In other words, in the 'live' condition the pattern of responses differed significantly from that in the 'die' condition (i.e. a framing effect was observed) whereas there was no evidence of a framing effect between the 'be saved' and 'die' conditions. This was a slightly surprising result,

⁷⁵ Again, these tell us whether the proportion of participants choosing each option in each framing condition differed significantly from the 'expected' proportion, where these 'expected' proportions are generated by assuming that each condition follows the overall distribution (in which 66.3% of respondents favoured the sure option, and 33.7% favoured the risky option).

since the classic ADP framing effect was not replicated (although the trend was in the expected direction).

6.3. Discussion

As in the first experiment, a framing effect was observed: in the ‘die’ condition, preferences shifted towards the risky option as compared with the other two conditions. However, in the second experiment, this effect was driven mainly by the ‘live’ and ‘die’ conditions, *not* the ‘be saved’ condition. Therefore, the results of the second experiment were contrary to the deflationary hypothesis, which predicts that the framing effect should be eliminated (or attenuated) by replacing ‘be saved’ with ‘live’. Accordingly, that hypothesis is not supported by the new data presented here. I conclude that it is unlikely that the ADP framing effect depends on ‘be saved’ and ‘die’ failing to conventionally denote jointly exhaustive sets.

To round up the discussion of the two experiments, there is little evidence that the framing effect produced by the ADP can be explained away in a deflationary manner, merely by appeal to a logical asymmetry between the framing conditions, resulting from the predicates ‘be saved’ and ‘die’ failing to be contradictory. In the remainder of this chapter, I broaden my scope to consider how other framing paradigms escape the deflationary challenge.

7. Exhaustive Alternatives

Compounding the lack of direct support for a deflationary account of the ADP framing effect, other framing studies have used pairs of predicate expressions that are more plausibly contradictory. I switch my focus here from risky-choice framing (given that the ADP casts its shadow over so much of that literature) to attribute framing (which also has the benefit of being rather a more straightforward paradigm). In this section, I present a couple of examples of attribute framing studies that use genuinely contradictory predicate expressions.

First, in a study by Wilson et al. (1987) participants are presented with a hypothetical scenario in which they are offered a risky surgery to treat terminal liver disease. In the event that the surgery does not prove fatal, it will extend life expectancy from one year to five years. In one framing condition, the outcome of the surgery is described in terms of the probability of ‘surviving’ the operation (at various probability levels of 10%, 20%, 40%, 60% and 80%). In the other framing condition, the outcome is described in terms of the probability of ‘dying’ during the operation (at the complementary levels of 90%, 80%, 60%, 40% and 20%). Participants are then asked whether or not they would opt for surgery.

In line with the standard ‘valence-consistent’ shift, the researchers find that participants in the positively-worded (‘surviving’) conditions are significantly more likely to opt for surgery than

those in the negatively-worded ('dying') conditions. This is despite the fact that 'surviving' and 'dying' here plausibly denote sets that are mutually exclusive (since one cannot both survive an operation and die during it) and jointly exhaustive (one either survives or dies, with no third possibility). In this example, the framing effect cannot depend on inexhaustive alternatives.

In a more recent study (briefly discussed in chapter 1), Leong et al. (2017) ask participants to imagine that they are recruiting players for a college basketball team. In one framing condition a player is described in terms of the proportion of shots he 'made' last season (at probability levels of 40% or 60%). In another, the player is described in terms of the proportion of shots he 'missed' (at the complementary levels of 60% or 40%).

Again, in line with the standard attribute framing effect, participants in the positively-worded ('made') condition rated the player as more valuable to their basketball team than those in the negatively-worded ('missed') frame. Here again, 'made' and 'missed' seem genuinely contradictory, being mutually exclusive (since shots that are made are not missed, and vice-versa) and jointly exhaustive (shots are either made or missed, with no third possibility). Accordingly, the sentences do not suffer from the kind of semantic asymmetry that afflicts the ADP.

These examples (and several others like them, some of which will be discussed at later points in the thesis) show that the deflationary challenge is not applicable to all experiments in which framing effects are observed. Crucially, then, even if we could dissolve the particular puzzle of the ADP framing effect (notwithstanding the evidence presented in the previous two sections), we cannot dissolve the general puzzle of framing effects. Wider empirical research points to the presence of a substantive phenomenon that remains in need of explanation. In subsequent chapters I will consider what the explanation might be. For now, though, I want to consider a potential objection to the argument I have presented.

8. Conventional Meaning and Representation

One might worry that, even where alternative frames use predicates with genuinely contradictory conventional meanings, hearers may still fail to *represent* them as such. Instead, the representations they form might be compatible with the denoted sets overlapping (thus failing to be mutually exclusive) or omitting some entities in the domain (thus failing to be jointly exhaustive). The idea here is reminiscent of classic 'Frege puzzles' in which, for example, the meaning of 'Hesperus' may be represented in a way that would be compatible with its having a different denotation from 'Phosphorus', even though both expressions actually denote the same celestial body. I will pursue the analogy a little further in the next subsection.

8.1. Fregean Sense and Reference

Frege (1948) famously argued for a distinction between two components of the meaning of a linguistic expression: its sense ('Sinn') and its reference, or extension ('Bedeutung'). He was led to do so by considering pairs of identity statements, which are extensionally identical but appear to have different epistemic implications. For example, although a true statement of 'a=b' has the same truth conditions as 'a=a', 'a=b' seems capable of being informative in the way that 'a=a' is not. Moreover, this difference appears to have truth-conditional implications when co-extensional expressions are substituted for one another in certain syntactic environments. For example, at a certain point in history, many astronomers would not have assented to the claim, 'Hesperus is Phosphorus', despite the fact that they would (presumably) have assented to the claim, 'Hesperus is Hesperus'. Intuitively, then, (5), below, expresses something true of these astronomers, while (6) expresses something false of them.

(5) The astronomers believed that Hesperus was Hesperus.

(6) The astronomers believed that Hesperus was Phosphorus.

However, both sentences identify one and the same entity with itself, since 'Hesperus' and 'Phosphorus' share a single extension (unbeknownst to the astronomers). It seems, then, that 'Hesperus' and 'Phosphorus' must contribute something other than their extensions to the truth-conditional meanings of sentences like (5) and (6). Frege was moved to posit 'senses' to play that role. Precisely what Fregean senses are supposed to be is a question that has occupied philosophers of language throughout the last century. In the next section, I will appeal to some of their proposals in arguing that the notion of sense will not help explain framing effects.

First, though, it is important to note a disparity between framing puzzles and Frege puzzles: pairs of expressions like 'survive' and 'die', or 'made' and 'missed', do not have *identical* denotations. In this respect, they are unlike 'Hesperus' and 'Phosphorus'. Rather, their denotations are mutually exclusive and jointly exhaustive. Nevertheless, the expressions might still be thought to have Fregean senses, which mediate – and potentially obscure – the relationship between their denotations. Therefore, the disparity does not prevent us from hypothesising that what is being observed in the psychology lab is a kind of Frege-puzzle. Although this is an initially intriguing idea, I will argue that it cannot be the case: whether or not such expressions have Fregean senses, such senses would remain irrelevant to the representational differences that mediate framing effects.

8.2. Why Framing Puzzles are not Frege Puzzles

Recall, first, that pairs of predicate expressions like ‘survive’ and ‘die’, or ‘made’ and ‘missed’ do not just happen to denote mutually exclusive and jointly exhaustive sets in the actual world but in *all possible worlds*. This shows that the sense-reference distinction can be of no use if it is thought of as a distinction between extensions and intensions. That approach might resolve some Frege-type cases (although probably not the ‘Hesperus’/ ‘Phosphorus’ case). For example, although ‘Boris Johnson’ and ‘The UK prime minister’ do, in fact, share an extension at the time of writing, there are possible worlds in which the current UK prime minister is someone else. In contrast, ‘Boris Johnson’ refers to the same individual in *all* of these possible worlds. Therefore, there are some possible worlds in which the identity does not hold, meaning that ‘Boris Johnson’ and ‘The UK prime minister’ differ in their intensions. In contrast, there are no worlds in which someone might both survive and die, or neither survive nor die (nor in which a shot might be both made and missed, or neither made nor missed). Therefore, these expressions have contradictory intensions as well as contradictory extensions. Appealing to Fregean sense would be futile, then, if sense were thought of merely as an expression’s intension.

Moreover, the contradictory intensions of predicate expressions like ‘survive’ and ‘die’, or ‘made’ and ‘missed’, depend on real-world properties that are mutually exclusive and jointly exhaustive. Therefore, it will not help to appeal to *hyperintensions* either. By way of contrast, consider the pair of predicate expressions, ‘cordate’ and ‘renate’, building on the discussion of Quine (1951). Although these are applicable to exactly the same entities in the actual world (and perhaps also in all possible worlds, if having a heart and having kidneys are necessary concomitants), the predicates concern *distinct* properties: namely, having a heart vs. having kidneys. In other words, the expressions depend in a fine-grained way on different real-world properties. In contrast, pairs of expressions like ‘survive’ and ‘die’, or ‘made’ and ‘missed’, have mutually exclusive and jointly exhaustive denotations in virtue of real-world properties that are themselves mutually exclusive and jointly exhaustive. Therefore, appealing to Fregean senses will not help if the sense of an expression is thought of as its hyperintension.⁷⁶

Evans cashes out the notion of sense by appealing to Frege’s ‘Intuitive Criterion of Difference’, which he characterises as follows:

[T]he thought associated with one sentence, *S* as its sense must be different from the thought associated with another sentence *S'* as *its* sense, if it is possible for someone to

⁷⁶ For similar reasons, we can set aside Dummett’s (1978, 1981) conception of sense, which takes differences in sense to be manifested in differences in the procedures for establishing real-world extensions. Whether we are talking in terms of ‘surviving’ or ‘dying’, ‘making’ or ‘missing’, the procedures would seem to be the same.

understand both sentences at a given time while coherently taking different attitudes towards them, i.e. accepting (rejecting) one while rejecting (accepting), or being agnostic about, the other. (Evans, 1982, pp. 18-19, emphasis in original)

This criterion doesn't apply to frames either. Importantly, framing effects do not trade on a mere lack of knowledge on the part of hearers. It would clearly be unreasonable to expect people to represent the meanings of 'survive' and 'die', or 'made' and 'missed' as contradictory if they did not know them to be so (just as it seems unreasonable to expect people to represent 'Hesperus' and 'Phosphorus', or 'cordate' and 'renate' as co-extensional if they do not know they are). However, I submit that ordinary speakers of English *do* know that 'survive' and 'die', or 'made' and 'missed' have contradictory meanings. Framing cases are therefore quite unlike Frege cases where, for example, the astronomers simply don't know that 'Hesperus' and 'Phosphorus' denote the same entity.⁷⁷ Arguably, understanding the meanings of 'survive' and 'die', or 'made' and 'missed' requires understanding them to be contradictory. Therefore, if distinct senses are posited in accordance with the intuitive criterion of difference, such that competent speakers can fail to know the relation between the two expressions' denotations, they will be irrelevant to framing cases.⁷⁸

Incidentally, it is worth noting that hearers of frames are not expected to be logically omniscient. They are not required to draw a logical inference from, for example, the fact that a subset of individuals will die, to the fact that the remainder will survive. Rather, it should be sufficient that they represent one or other of those subsets. Since, in fact, the subsets *are* mutually exclusive and jointly exhaustive (and this is required by the conventional meanings of the predicates), the issue of which subset is represented cannot make a difference to one's subsequent judgements. In effect, the point is not that $x\%$ of people surviving *entails* $100-x\%$ dying; the point is that this is one and the same thing.

The preceding discussion suggests that, when hearers form distinct representations in response to alternative frames, these representations cannot be traced back purely to the conventional meanings of the frames (which generate logically equivalent representations) but depend on wider factors. This is an idea I pick up in the second half of the thesis.⁷⁹

⁷⁷ Bermúdez (2018) similarly points out that framing cases do not depend on mere ignorance. I will consider his positive proposal below.

⁷⁸ Relatedly, we can set aside the conception of sense endorsed by Taschek (1992). Taschek understands senses as potentials for reorganising a thinker's beliefs about entities. However, since the thinkers in question already know that expressions like 'survive' and 'die', or 'made' and 'missed' have contradictory conventional meanings, there is no reason to posit senses that obscure those facts.

⁷⁹ The idea may be compatible with *pragmatic* interpretations of Fregean sense. For example, Bezuidenhout (1997) thinks of senses as context-dependent modes of presentation rather than as constant aspects of an expression's meaning. This chapter has been concerned with whether framing

In sum, although, by assumption, alternative frames *are* represented differently, I do not believe the relevant representational differences can be traced back to the conventional meanings of predicate expressions that are substituted across framing conditions – regardless of whether we focus on sense or reference. There is every reason to think that hearers represent these expressions as having genuinely contradictory conventional meanings. Before concluding, though, I want to briefly consider a proposal currently being developed by Bermúdez.

8.3. Ultraintensionality

Bermúdez (2009, 2018) suggests that, even when alternative frames are known to describe the same outcome, hearers can sometimes be rational to respond to them differently. He proposes that the scenarios used in framing studies (and, more broadly, in valuation tasks) establish ‘ultraintensional contexts’. These are related to intensional and hyperintensional contexts in the following way:

The defining feature of an intensional context is that it allows for (truth- and/ or rationality-preserving) failure of substitution of materially equivalent sentences. By extension, a *hyperintensional* context allows for (truth- and/ or rationality-preserving) failure of substitution of logically equivalent sentences... An ultraintensional context is a context that allows for (truth- and/ or rationality-preserving) failure of substitution of *known* identities. (Bermúdez, 2018, p. 184)

The idea, then, is that alternative frames, in spite of their being known to describe the same state of affairs, can make different features of a situation more readily available to reasoners; and this licenses systematically different responses. Bermúdez writes:

In sum, there is a range of frame-sensitive factors that are rationally relevant to how an agent values outcomes. One might reasonably expect them to emerge more clearly on some frames than on others. In particular, it seems perfectly reasonable to think that they might rationally attach to an outcome when it is framed one way, but not when it is framed another way – even when the agent is aware that the two frames correspond to a single outcome. If all that is true, then valuation is an ultraintensional context. (Bermúdez, 2018, p. 202)

effects can be traced back to the *conventional* meanings of predicate expressions substituted across framing conditions (and thus deflated or dissolved). Therefore, I have not yet addressed pragmatic analyses, which take framing effects seriously as substantive puzzles to be resolved.

Bermúdez does not say much about exactly what it is that makes certain factors ‘frame-sensitive’, i.e. why frames should induce different representations of a situation.⁸⁰ This is the question with which I am primarily concerned in the thesis. I believe that the explanations I will explore in forthcoming chapters can circumvent the appeal to ultraintensional contexts altogether. I therefore suggest that this may prove to be the better path to follow.

9. Conclusion

In this chapter, I have considered – and rejected – one attempt to dissolve the puzzle of framing effects. The deflationary challenge I have addressed appeals to the semantic inequivalence of alternative frames used in the classic ‘Asian Disease Problem’, which results from the conventional meanings of the substituted predicates (‘be saved’ and ‘die’) failing to have jointly exhaustive denotations. I have shown that this is unlikely to explain the framing effect associated with the ADP, and that the deflationary challenge does not extend to other framing paradigms. In the next chapter, I shift my focus to consider whether the number expressions used in alternative frames are genuinely *complementary*.

⁸⁰ In personal communication, he confirms that this issue will be covered in more detail in his forthcoming book *The Power of Frames*.

CHAPTER 4

Lower-Boundedness

Abstract

In this chapter, I focus on the interpretation of number expressions that appear in alternative frames. Specifically, I investigate whether framing effects could result from these expressions being assigned lower-bounded ('at least') readings. Based on my analysis of recent experimental data, both from the framing literature and wider psycholinguistic research, I argue that this could well be a contributory factor (although it seems unlikely to explain framing effects in their entirety). In parallel, I argue that there is relatively limited support for a competing account of framing effects, based on 'fuzzy trace theory'. I end by discussing whether lower-boundedness should be thought of as a semantic or a pragmatic phenomenon, and what that implies for a rationalising explanation of framing effects.

1. Introduction

Equivalency frames standardly include number expressions. Consider again the 'Asian Disease Problem' (ADP), reproduced below:

Imagine the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume the exact scientific estimates of the consequences are as follows:

Framing condition 1:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.

Framing condition 2:

If Program C is adopted, 400 people will die.

If Program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.

As in the previous chapter, I reorganise the four sentences that appear in the framing conditions, to pair up (1) with (2), and (3) with (4):

(1) If Program A is adopted, 200 people will be saved.

(2) If Program C is adopted, 400 people will die.

(3) If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.

(4) If Program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.

Each of these sentences includes one or more number expressions, used to denote subsets of the 600 people expected to be killed by the disease.⁸¹ As mentioned in the previous chapter, the logical equivalence of each pair of frames – (1) and (2), (3) and (4) – depends on the number expressions in the corresponding clauses having *complementary* meanings. So, for example, the quantities denoted by ‘200’ in (1) and ‘400’ in (2) are supposed to sum to 600 (the number of people expected to be killed by the disease).

In this chapter, I will challenge the complementarity assumption. I will argue that framing effects depend, at least in part, on number expressions being interpreted *lower-boundedly*: for example, ‘200’ would be represented as denoting *at least 200*, and ‘400’ as *at least 400*.⁸² The structure of the chapter is as follows: in §2 I introduce the notion of a ‘scalar expression’. In §3 I describe the assumption, standardly made in the framing literature, that the number expressions in alternative frames have ‘bilateral’ (‘exactly’) meanings. In §4 I discuss experiments that suggest the framing effects in the ADP depend on lower-bounded (‘at least’) interpretations of number expressions. In §5 I discuss some counterevidence, and a competing account of framing effects based on ‘fuzzy trace theory’. In §6 I argue that a modest version of the lower-bounding hypothesis is better supported than the fuzzy trace account. In §7, I describe a spectrum of approaches to the semantics and pragmatics of number expressions, and indicate how each could support a rationalising explanation of framing effects.

2. Scalar Expressions

Number expressions are a species of ‘scalar’ expression. Scalar expressions (or ‘scalars’) can be thought of as denoting points or regions along a linear ordering (or ‘scale’). Numbers form one such scale, points along which are denoted by number expressions. Later in the chapter, I will be concerned with other kinds of scalar expression too, specifically the existential quantifier, ‘some’, and the universal quantifier, ‘all’.

Scalars have received attention from many philosophers and linguists, prominently Horn (1989). It is well-established that these expressions can generate different interpretations in

⁸¹ Frames (3) and (4) also include number expressions in the probabilities. I will largely ignore these, aside from a brief discussion in §5.

⁸² My focus in the chapter will be mainly on *quantifier* uses of number expressions. However, I will briefly discuss predicative uses in §7, in describing the semantic analysis developed by Geurts (2006).

different contexts, denoting regions on a scale that are lower-bounded, upper-bounded, or bounded on both sides. Take, for example, an utterance of sentence (5) below:

(5) With these 12 letters it is possible to make 600 words.

Imagine that the speaker continues by uttering (5a):

(5a) Anyone who identifies 200 words can go through to the next round.

In (5a), '200' is most naturally read lower-boundedly, as denoting *at least 200*. After all, someone who identified 300 words could presumably go through to the next round too (whereas someone who identified 100 could not).

In other possible continuations, such as (5b), '200' is naturally read *bilaterally*, as denoting *exactly 200*:⁸³

(5b) The world record holder identified 200 words in two minutes.

On this interpretation, the denotation of '200' is taken to have both a lower bound and an upper bound. Somewhat less commonly, scalar expressions can also generate purely upper-bounded readings, as in the continuation in (5c), where '200' is naturally read as denoting *at most 200*:

(5c) There will only be time to find 200.

It remains controversial which, if any, of these denotations corresponds to number expressions' conventional meanings, and which are the result of pragmatic enrichment. I return to that question in §7, having first argued that framing effects may be driven, at least in part, by number expressions being *interpreted* lower-boundedly. This fact stands in contrast to an assumption, standardly made by framing researchers, that they should be interpreted bilaterally.

3. The Bilateral Assumption

In order for the number expressions appearing in (corresponding clauses of the) alternative ADP frames to have complementary conventional meanings, these meanings must be bilateral and precise. For example, '200' in (1) must denote *exactly 200*, and '400' in (2) must denote *exactly 400*, in order for their denotations to sum to *exactly 600* (the number of people expected to be killed by the disease).⁸⁴ Accordingly, the logical equivalence of the alternative frames

⁸³ Or, perhaps, *approximately 200*. This would still be a bilateral reading, just a less precise one.

⁸⁴ Tversky and Kahneman may have been attempting to secure this reading by referring to 'exact scientific estimates' in the vignette. However, it is not obvious that this is successful and, as far as I am aware, they do not check whether it is.

depends on the number expressions having bilateral meanings. Nevertheless, it has been suggested that they are often interpreted unilaterally – specifically, lower-boundedly.

If the number expressions appearing in frames were interpreted lower-boundedly, then they would no longer be understood as having complementary denotations. Specifically, ‘200’ in (1) would be interpreted as denoting *at least 200*, while ‘400’ in (2) would be interpreted as denoting *at least 400*. Taken together, these denotations need not sum to exactly 600.⁸⁵ Crucially, that could explain the standard framing effect. This is because (1) would be interpreted as in (1a) and (2) as in (2a):

(1a) If Program A is adopted, at least 200 people will be saved.

(2a) If Program C is adopted, at least 400 people will die.

It is consistent with (1a) that more than 200 people will be saved under Program A, whereas (2a) ensures that no more than 200 people can be saved under Program C. Therefore, Program A could have a better outcome than Program C. If number expressions were being interpreted lower-boundedly, that could explain the shift towards the sure option in the ‘be saved’ framing condition, and away from the sure option in the ‘die’ condition.⁸⁶ According to the lower-bounding hypothesis, then, the ADP framing effect depends on the number expressions in the sure options being assigned ‘at least’ readings.⁸⁷ Hypotheses in this ballpark have been put forward on a number of occasions since the early days of framing research, including by Macdonald (1986). However, it is only recently that empirical investigations have been conducted.

4. Mandel’s Experiments

I will first describe three experiments reported by Mandel (2014). To give an overview of the results, he was able to eliminate framing effects entirely by encouraging bilateral readings of the number expressions. He also found evidence that framing effects are facilitated by lower-bounded readings. Since the details will be important, I will describe the experiments in some depth.

⁸⁵ I assume that the number expressions in (3) and (4) are straightforwardly complementary: since ‘600’ exhausts the number of people domain, it cannot be read as ‘at least 600’.

⁸⁶ In contrast, if hearers formed *upper*-bounded representations of number expressions, we would predict a *reversal* of the standard framing effect.

⁸⁷ It remains to be explained why alternative frames are typically represented as being equivalent when *both* are presented. Perhaps it might be argued that bilateral interpretations of the number expressions become more readily available in such contexts, as compared to when only one framing condition is presented. This would be an interesting hypothesis to investigate but I cannot pursue it further here.

Mandel's first experiment deployed the ADP in a within-subjects design, such that participants received *both* framing conditions. The order in which the two framing conditions were presented was counterbalanced between subjects, and all participants received a series of unrelated, intervening tasks. For one group of participants, the sure options in each framing condition were presented with the modifier 'exactly' appearing before the number expression, as in (1b) and (2b) below. Mandel takes this to secure the co-extensionality of the framing conditions.⁸⁸

(1b) If Program A is adopted, exactly 200 people will be saved.

(2b) If Program C is adopted, exactly 400 people will die.

For the remaining participants, the sure options contained the modifier 'at least' before the number expression, as in (1a) and (2a) above.

Before reporting the results of this experiment, I will describe Mandel's closely related second experiment. This experiment deployed a similar risky-choice paradigm but used a between-subjects design.⁸⁹ As in the first experiment, participants were divided into 'at least' and 'exactly' conditions. The second experiment also included a pair of conditions in which no modifier was included in the sure options, thus mirroring the classic ADP.

In line with the lower-bounding hypothesis, no framing effects were observed in the 'exactly' conditions, in either experiment. In other words, once the modifier 'exactly' was added, no statistically significant difference was observed between the choices made in the 'be saved' and 'die' conditions. Specifically, in the first experiment, most participants in the 'exactly' condition (73.3%) chose the same option in both framing conditions; there was therefore no within-

⁸⁸ As discussed in the previous chapter, that is not strictly right, since 'be saved' and 'die' are imperfect antonyms. As will be discussed, Mandel found no framing effects across the 'exactly' conditions. This provides some further evidence against the deflationary hypothesis discussed in the previous chapter (which predicts that framing effects should still be observed in the 'exactly' conditions).

⁸⁹ The modified scenario, taken from (Mandel, 2001), was intended to make it more plausible that the number of lives under threat could be accurately forecast. The scenario reads:

In a war-torn region, the lives of 600 stranded people are at stake. Two response plans with the following outcomes have been proposed. Assume that the estimates provided are accurate. The options are then phrased similarly to those in the ADP, except that: (i) the sure options include the phrase 'it is certain that' before the number expressions; (ii) the risky options remove reference to 'people', instead referring either to 'all 600' or 'nobody'; (iii) in the risky option of the 'die' framing condition, the order of the complements is switched; and (iv) each of the options refers to 'Plans' rather than 'Programs'. Additionally, Mandel uses the following wording in the question to participants: 'Which of the two plans would you choose – A or B?' whereas the original ADP – and, it seems, Mandel's first experiment – asked 'Which of the two programs would you favour?'. I think it is unlikely that these changes are responsible for the pattern of results Mandel obtains, most notably the elimination of framing effects in the 'exactly' conditions (particularly since framing effects *were* observed in the other conditions that used the scenario).

subjects framing effect.⁹⁰ Meanwhile, in the second experiment, adding ‘exactly’ eliminated the standard between-subjects framing effect. The sure option was chosen by 59% of participants in the ‘exactly 200 people will be saved’ condition and by 43.2% in the ‘exactly 400 people will die’ condition. This difference was not statistically significant.

Conversely, the ‘at least’ conditions led to large framing effects in the standard direction. In the first experiment, 67.7% of participants in the ‘at least’ group preferred the sure option in the ‘be saved’ framing condition, and the risky option in the ‘die’ condition. In the second experiment, 91.7% of participants in the ‘at least 200 people will be saved’ condition preferred the sure option, compared with 32.5% of those in the ‘at least 400 people will die’ condition. This was a statistically significant difference.

The no-modifier conditions in the second experiment yielded a medium-sized framing effect: 57.9% of participants in the ‘be saved’ condition preferred the sure option, compared with 26.3% of participants in the ‘die’ condition. This is consistent with the hypothesis that some participants spontaneously adopt lower-bounded readings of the number expressions in the sure options, when no modifier is present.

In his final experiment, Mandel probed participants’ interpretations of the number expressions by manipulating what he calls the ‘explication’ of the sure and risky options. He notes that the sure options in the classic ADP are only ‘partially explicated’, leaving implicit the fate of the remaining 400 people (in the ‘be saved’ condition) or 200 people (in the ‘die’ condition).⁹¹

Mandel formulates ‘fully explicated’ versions of the sure options as follows:

(1c) If Plan A is adopted, it is certain that 200 people will be saved and 400 people will not be saved.

(2c) If Plan C is adopted, it is certain that 400 people will die and 200 people will not die.

Correspondingly, he formulates *partially*-explicated versions of the risky options, as follows:

(3a) If Plan B is adopted, there is a 1/3 probability that all 600 will be saved.

(4a) If Plan D is adopted, there is a 2/3 probability that all 600 will die.

⁹⁰ Mandel does not report on order effects, which might have indicated that some kind of framing effect was still operating, i.e. if participants who initially received the ‘be saved’ framing condition tended to choose the sure option in both conditions; whereas those who initially received the ‘die’ framing condition tended to choose the risky option in both conditions.

⁹¹ The observation that this may be driving the framing effect can be traced back at least to Kühberger (1995).

Combinations of fully- and partially-explicated options were deployed in a between-subjects experiment, with participants being divided between the six conditions represented in the rows of the table below:

| Framing condition | Explication of sure option | Explication of risky option |
|-------------------|----------------------------|-----------------------------|
| 'Be saved' | Partial | Full |
| | Full | Full |
| | Full | Partial |
| 'Die' | Partial | Full |
| | Full | Full |
| | Full | Partial |

Following the usual choice task, Mandel asked recipients in 'be saved' conditions the following questions (the versions used in the 'die' conditions are shown in square brackets):

1. Did you interpret Plan A to mean (a) at most, (b) exactly, or (c) at least 200 will be saved [400 will die]?
2. Did you interpret Plan B to mean there was (a) at most, (b) exactly, or (c) at least a 1/3 probability that all 600 people will be saved [2/3 probability that all 600 people will die]?

The results of this experiment show, first, an 'explication effect' on *choice*.⁹² A standard framing effect arose from the classic ADP conditions, in which the sure options were partially-explicated and the risky options were fully explicated (rows 1 and 4 in the table above). Specifically, 68% of participants in the 'be saved' condition chose the sure option, compared with 20% in the 'die' condition. In contrast, the framing effect was *eliminated* when both options were fully explicated (rows 2 and 5 in the table above): 59.1% of participants in the 'be saved' condition chose the sure option, compared with 50.0% in the 'die' condition (a non-significant difference). Responses even trended in the direction of a *reverse* framing effect when the sure options were fully explicated and the risky options were partially explicated (rows 3 and 6 in the table above): 42.3% of participants in the 'be saved' condition chose the sure option, compared with 68.0% in the 'die' condition. However, this effect fell short of statistical significance.

In addition to the effect on choice, Mandel found an effect of explication on the interpretation of number expressions. Partial explication of the sure options was correlated with participants

⁹² They thus provide some further support for the idea that classic risky-choice framing effects are driven by the sure option being only partially explicated. See also (Kühberger & Tanner, 2010).

reporting lower-bounded, ‘at least’ readings of the number expressions (in both the ‘be saved’ and ‘die’ conditions). Specifically, 64% of participants receiving partially-explicated sure options reported that they interpreted the number expressions lower-boundedly. Conversely, full explication of the sure options was correlated with participants reporting bilateral, ‘exactly’ readings of the number expressions: 62% of those who received the fully-explicated versions of the sure options reported adopting bilateral readings.⁹³

There was also an interaction between the choice patterns and the interpretations of the number expressions: an overall framing effect was observed for the sets of participants reporting lower-bounded readings of the number expressions in the sure options. Conversely, no framing effect was observed for the sets of participants reporting bilateral readings. A reverse framing effect was observed for the small sets of participants reporting *upper*-bounded readings.⁹⁴

To summarise Mandel’s results, framing effects were eliminated under bilateral interpretations of the number expressions in the sure options, but the effects re-emerged under lower-bounded interpretations. This suggests that, in classic risky-choice framing experiments, the results may be driven by many of the participants interpreting number expressions lower-boundedly.

Mandel concludes:

In short, the findings indicate that, for most people, Options A and C in the ADP are descriptions of different events, and not merely re-descriptions of the same event. Although effects of the usual ADP positive-negative manipulation are description effects, strictly speaking, they are not framing effects, and thus should not be labeled as such. (Mandel, 2014, p. 1193)⁹⁵

⁹³ With respect to the risky options, most participants reported bilateral interpretations of the probabilities, regardless of explication, although larger minorities reported lower-bounded interpretations when the risky option was only partially explicated (37%) compared with when it was fully-explicated (22%) (this was a statistically significant difference). For now, I set aside the interpretation of the probabilities but I return to this point briefly in §5.

⁹⁴ There was also some evidence of the converse pattern emerging in relation to the interpretations of the probabilities in the risky options. A framing effect in the standard direction was observed for the sets of participants reporting upper-bounded readings of the probability fraction. However, the reverse framing effect for the sets of participants reporting lower-bounded readings did not reach statistical significance.

⁹⁵ Mandel may not be right to conclude that lower-bounded interpretations would show the framing effect to be dissolved. As discussed in chapter 1, framing effects are *standardly* taken to depend on hearers forming distinct representations of states of affairs described by alternative frames (albeit this is usually thought to be triggered by the positive or negative wording used, rather than the number expressions). The key question is whether hearers are *licensed* to form distinct representations. I do not believe Mandel has shown that they are. In §7 I will discuss some possible arguments to that effect, which would trace lower-bounded interpretations to the conventional meanings of number expressions, or pragmatic enrichments of these. Although he cites the associated literature, Mandel explicitly defers discussion of it, writing:

Although Mandel's findings are striking, it may be a little premature to draw firm conclusions from them. Others have failed to replicate his experiments. Two subsequent studies, by Simmons and Nelson (2013) and Chick et al. (2016), present contradictory findings.⁹⁶ I explore the latter study next.

5. Chick et al.'s Experiments

Chick et al. (2016) present results from two experiments conducted in response to Mandel's study. They observe a methodological flaw in Mandel's third experiment: participants only reported their interpretations of the number expressions *after* completing the choice task. It is therefore possible that they were offering a post-hoc justification of their earlier response, rather than giving a reliable account of how they initially understood the options. In other words, it is plausible that the causation was running in the wrong direction: reports of lower-bounded interpretations might have been *caused by* the presence of a framing effect, rather than *causing* it. Chick et al. address this issue in their own experiments, by providing participants with comprehensive instructions to adopt bilateral interpretations of the number expressions, and probing these both before and after the choice tasks are completed. In the next two subsections I will describe each experiment in turn.

5.1. Experiment 1

In the first experiment, participants were initially trained on how to interpret partially-explicated options in ADP-style risky-choice framing scenarios. They were told that the numbers of people who will 'be saved' or 'die' were exact and would not increase or decrease. It was also explained that the probabilities in the risky options would always be complementary (adding to 1) and would range over all-or-nothing alternatives: for example, either no people would be saved or all would be. The clauses which specify these alternatives are termed the 'zero complement' and the 'non-zero complement', respectively. An 'initial ambiguity questionnaire' tested whether participants were interpreting the stated information as instructed. The questionnaire presented a single framing problem, either in the positively-worded or negatively-worded condition. Each of the sure and risky options were only partially explicated. Participants were then asked three questions: two relating to the interpretation of

Although a specification of how various types of context (e.g., conversational, semantic, and sentential) affect the interpretation of [numeric quantifiers] is beyond the aims of this article, the present research...suggests the need for a comprehensive psycholinguistic account of numeric quantifier use and interpretation. (Mandel, 2014, p. 1196)

Ultimately, the lower-bounding hypothesis cannot provide a *rationalising* explanation of framing effects until its linguistic and philosophical foundations have been further developed.

⁹⁶ See (Mandel, 2013) for a response to Simmons and Nelson, which points to some methodological and interpretational problems with their attempted replication.

the number expression in the sure option; and one relating to the interpretation of the probability in the risky option. An example is provided below (with the 'correct' interpretations in bold):

1. Did you assume that more than 300 people would be saved in option A?

Yes/ **No**

2. Which of the following did you assume about option A?

a) Exactly 300 saved

b) At least 300 saved

c) Some of the other 200 saved

3. Which of the following did you assume about option B?

a) 2/5 probability some saved

b) 2/5 probability all saved

c) 2/5 probability none saved

After responding, participants were given the correct responses, together with detailed explanations that reiterated the earlier instructions. They also received an additional set of worked examples.

Following this initial training stage, the choice trials consisted of 60 framing problems relating to 20 different ADP-style risky choice scenarios (where either lives or money were at stake). Each problem required participants to make the usual choice between a sure option and a risky option. The set of 60 framing problems received by each participant included an equal split of positively- and negatively-worded conditions. The sure options were partially explicated throughout, whereas risky options were manipulated to present (with equal frequency) either the zero complement only, the non-zero complement only, or both complements. No participant received both the positively- and negatively-worded conditions of a given problem with the same manipulation of the risky option. After each response, participants rated their confidence in their decision on a five-point scale. In the final phase of the experiment, a post-framing ambiguity questionnaire re-tested participants' interpretation of partially-explicated options.

Participants were removed from the analysis if they answered any of the questions on either ambiguity questionnaire incorrectly. This reduced the number of participants from 81 to 41.⁹⁷ The crucial result was the observation of a standard framing effect, despite the participants apparently adopting bilateral interpretations of the number expressions. Participants' confidence ratings also tracked the framing effect: the sure option was chosen with greater confidence in the positively-worded frame, and the risky option with greater confidence in the negatively-worded frame.⁹⁸ The results of the first experiment suggest that lower-bounded interpretations are not necessary for risky-choice framing effects to occur, *pace* Mandel.⁹⁹

5.2. Experiment 2

The second experiment by Chick et al. used a similar procedure, with an initial training stage and pre- and post-framing questionnaires. However, participants received only a single framing problem rather than 60. In an attempt to replicate as closely as possible the 'exactly' condition of Mandel's second experiment, the experimenters used the same scenario and included the modifier 'exactly' in each of the sure options. For each participant, the target trial was administered in the same (positively- or negatively-worded) framing condition as the problem in the preceding ambiguity questionnaire. And each participant received one of the three manipulations of the risky option that were used in Chick et al.'s first experiment. In addition, time limits on decision making and confidence ratings were removed; and a question was also added into the ambiguity questionnaires that tested participants' ability to apply their interpretations of partially-explicated options in a deductive reasoning task.

⁹⁷ I return to discuss the high attrition rate in §5.4.

⁹⁸ Both effects were largest when only the zero complement of the risky option was presented, and were eliminated when only the non-zero complement was presented. I discuss this in §5.4.

⁹⁹ I take Mandel and Chick et al. to be making directly conflicting claims about whether lower-bounded interpretations are necessary for framing effects to arise in ADP-style risky-choice framing paradigms. However, Chick et al. sometimes write as though their claims are compatible with Mandel's, ascribing to him a mere *sufficiency* claim. For example:

Although Mandel's findings demonstrate that a lower-bounded interpretation is sufficient to produce framing effects, they do not demonstrate that such an interpretation is necessary to produce framing effects. (Chick et al., 2016, p. 240)

Later, they add:

In contrast to Mandel's argument about sufficiency, we make an argument about necessity: Our results are novel in demonstrating that linguistic ambiguity is not necessary to risky choice framing effects. (Chick et al., 2016, p. 251)

Yet, as the authors also note, "if disambiguation such as "exactly" eliminated framing effects, this would be evidence that ambiguity is necessary to framing" (Chick et al., 2016, p. 248). This, I take it, is precisely what Mandel finds with respect to a set of ADP-style risky-choice framing paradigms. As such, he *does* make the necessity claim that Chick et al. seek to refute. Indeed, Chick et al. acknowledge elsewhere that their results "contradict the conclusions reported by Mandel (2014)" (Chick et al., 2016, p. 251). The dialectic is therefore rather muddled, and the necessity/ sufficiency distinction seems to be something of a red herring. As stated, then, I understand both studies to concern the claim that lower-bounded interpretations of number expressions are necessary for framing effects to arise in certain ADP-style risky-choice framing paradigms.

Again, framing effects persisted among the 154 participants (of an original 291) who answered both ambiguity questionnaires entirely correctly. As in the first experiment, confidence ratings patterned with the observed framing effect.¹⁰⁰ The results of the second experiment provide further support for the authors' claim that lower-bounded interpretations of number expressions are unnecessary for framing effects to arise in ADP-style risky choice paradigms. Instead, such effects are also observed under bilateral interpretations.¹⁰¹ In the next subsection, I introduce the alternative account of framing effects that Chick et al. advocate.

5.3. Fuzzy Trace Theory

Fuzzy trace theory (henceforth FTT) was originally developed in the mid-1980s in the context of memory research (Brainerd & Kingma, 1984, 1985). Its central claim is that people simultaneously encode information in precise and 'gist-like' ways. Of most relevance for the current discussion, the proponents of FTT have provided an account of how numbers are represented during numerical reasoning tasks, including risky-choice framing paradigms (Chick et al., 2016; Reyna & Brainerd, 1991, 2011; Reyna, Chick, Corbin, & Hsia, 2014). They propose that multiple mental representations are activated in parallel, ranging along a continuum of precision, from verbatim representations (like 'exactly 200') to the most 'gist-like' representation, 'some'. During episodes of reasoning, thinkers are held to deploy the least precise representation that would be compatible with completing the task. It is predicted, then, that thinkers will operate with a representation of any positive number as 'some', whenever that enables sufficient differentiation between alternative possibilities.

Of course, 'some' will not always be sufficient: for example, if one wishes to establish whether it is better to have £200 or £300 then comparing 'some pounds' with 'some pounds' won't help. In such scenarios, thinkers must ascend the precision hierarchy. Proceeding stepwise, they are held to deploy incrementally sharper representations (for example, 'more' and 'less') in order to complete the task at hand. It is predicted that the maximally precise, verbatim representations will be used only in the last resort.¹⁰²

¹⁰⁰ Again, both effects were increased by presenting only the zero complements of the risky options, and eliminated by presenting only the non-zero complements. I return to this in §5.4.

¹⁰¹ In places, Chick et al. seem to suggest that their results refute *all* 'linguistic ambiguity' hypotheses, which propose that risky-choice framing effects arise from the partial explication of the sure options. I do not believe they have shown this. Indeed, in chapter 9 I will suggest that the 'information leakage account', to be introduced in chapter 6 of the thesis, can be extended to risky-choice framing paradigms. I therefore restrict discussion here to the more specific lower-bounding hypothesis.

¹⁰² Note that this is different from merely hypothesising that number expressions are interpreted *approximately* (whereby, for example, '200' is represented as denoting *about 200*). That hypothesis would not necessarily get us any further towards explaining framing effects since the framing conditions would still be represented as being equivalent to one another (they would just be equivalently approximate). As

On FTT's analysis of the ADP, framing effects are explained in terms of the categorical contrast between 'some' people in the sure options and 'no' people in the zero complements of the risky options. Spelling this out, in the 'be saved' condition, FTT predicts that the sure option is represented as guaranteeing that *some* people will be saved. According to the risky option, though, there is *some* probability of *some* people being saved (the non-zero complement) and also *some* probability of *no* people being saved (the zero complement). Since an outcome in which some people are saved for sure is preferable to the possibility of no people being saved, FTT predicts that the sure option will typically be preferred in this framing condition.

The converse story can be told for the 'die' condition: under the sure option, *some* people will die. Under the risky option, there is *some* probability of *some* people dying but also *some* probability of *no* people dying. In this case, a possibility of no people dying is preferable to the certainty that some people will die, making the risky option preferable. Taken together, these hypotheses can potentially explain the classic ADP framing effect.

Before moving on, I want to raise a conceptual problem with FTT. As far as I am aware, its proponents nowhere provide an explicit semantics for the 'gist-like' representation 'some' – specifically, they do not say whether it is meant to have a lower-bounded meaning (denoting some and possibly all) or a bilateral one (denoting some but not all). This is unfortunate given the linguistic controversy that surrounds the meanings of scalar expressions, and the central role the term 'some' plays in the FTT account.

On closer inspection, it seems that the 'gist-like' representation, 'some', must be taken to have a *lower-bounded* meaning, denoting *some and possibly all*. Otherwise, if 'some' meant 'some but not all' the fuzzy trace theorists could not claim that the non-zero complements of the risky options, in which *all* 600 people will be saved/ will die, are represented as *some* people being saved/ dying.

Interestingly, it follows that the sure options in each framing condition become inequivalent again: in the 'be saved' frame, some *and possibly all* people will be saved; in the 'die' frame, some *and possibly all* will die. It will thus be difficult for FTT to maintain that participants do not naturally form lower-bounded representations of number expressions. Presumably, however, they would continue to deny that their doing so is causally responsible for framing effects. According to FTT, the framing effect produced by the ADP arises purely as a result of the categorical contrast between *some* people and *no* people.

we will see, the FTT account depends crucially on the stronger claim that number expressions are represented as denoting *some*.

A more nuanced position seems available here. On the one hand, the experimental evidence presented by Chick et al. does suggest that lower-bounded representations are not *necessary* for framing effects to arise. On the other hand, the lower-bounding hypothesis could still provide *part* of the explanation of framing effects. In the remainder of the chapter I present further arguments for this modest version of the lower-bounding hypothesis.

5.4. Weighing the Evidence

Given the failure by Chick et al. to replicate Mandel's results, it seems advisable to treat with caution the claim that risky-choice framing effects depend entirely on lower-bounded interpretations of number expressions. However, I will argue that such interpretations are at least likely to *contribute* to framing effects, even if they cannot explain them in full. It is notable, for example, that nearly half of Chick et al.'s participants still failed to adopt the 'correct' bilateral interpretations of the partially explicated options in the ambiguity questionnaires, despite having received comprehensive instructions and training (as discussed, the responses of these individuals were removed from the final analysis). The authors do not speculate about why this was. Plausibly, it may have proved difficult to dispel lower-bounded readings of the number expressions. If so, these readings can be expected to arise spontaneously in other framing studies where experimenters make little or no attempt to secure bilateral interpretations. And, where they do arise, it makes sense that they would contribute to framing effects.¹⁰³

That said, it must be acknowledged that FTT provides a nice explanation of why framing effects are eliminated or reversed when the *risky* options are only partially-explicated. As discussed, the theory places crucial importance on the contrast between the sure options and the *zero complements* of the risky options. In the classic ADP, these are the clauses which state that *no* people will be saved (in the 'be saved' frame); or that *no* people will die (in the 'die' frame). In contrast, the non-zero complements (which state that 600 people will be saved/ will die) are predicted by FTT to be represented as being on a par with the sure options, since *some* people will be saved/ will die either way. FTT thus predicts that framing effects will only be observed when zero complements are presented, either alone or together with the non-zero complement; and that the effects will be strongest when the zero complements are presented alone. This

¹⁰³ Indeed, in providing a gloss of Mandel's results, Chick et al. write:

Adding "at least" to the sure option makes the value of the sure option higher than that of the risky option in the gain frame, and lower than that of the risky option in the loss frame. Naturally, subjects choose the option with the higher value, resulting in most subjects choosing the sure option in the gain frame and the risky option in the loss frame. (Chick et al., 2016, p. 251)

By the same logic, subjects working with the gist-like representation 'some and possibly all' should choose the sure option in the 'be saved' condition and the risky option in the 'die' condition.

prediction is borne out by the findings of Chick et al. and a number of other studies: for a review, see (Kühberger & Tanner, 2010); see also (Reyna et al., 2014).

Although FTT provides a good account of these results, I believe the lower-bounding hypothesis can explain them too. After all, if the probabilities in the partially explicated risky options are represented as lower-bounded, then the zero complement in the ‘die’ frame is potentially better than the zero complement in the ‘be saved’ frame. This is because an *at least* one-third probability that no people will die is potentially better than an *at least* two-thirds probability that no people will be *saved*. The first is consistent with the probability of everyone being saved being greater than one third. In contrast, under the second, the probability of everyone being saved cannot exceed one third (and may be less than that). This could explain why the framing effect is strengthened when only the zero complements are presented: it is not only that the sure option in the ‘be saved’ condition is better than the sure option in the ‘die’ condition, but that the risky option in the ‘die’ condition is better than the risky option in the ‘be saved’ condition. In other words, participants would end up with two, mutually reinforcing reasons to err towards the sure option in the ‘be saved’ condition and the risky option in the ‘die’ condition.¹⁰⁴

By the same reasoning, when only the non-zero complement is presented, the incentives are less clear. The non-zero complement in the ‘be saved’ frame is potentially better than the non-zero complement in the ‘die’ frame: an *at least* two-thirds probability that all 600 will be saved is potentially better than an *at least* one-third probability that all 600 people will die. The first is consistent with a higher probability of everyone staying alive. This could explain why the framing effect is weakened when only the non-zero complements are presented: participants have less reason to favour different options under each condition.

Overall, then, while the lower-bounding hypothesis can be expected to explain, at least in part, the framing effects observed in ADP-style risky-choice framing experiments, it may not be necessary to appeal to FTT’s ‘gist-like’ representations in order to account for the data. In the next section, I will show how wider psycholinguistic research lends further credence to the lower-bounding hypothesis over FTT.

6. Precise Representations

As discussed in §2, it is well understood that number expressions (among other scalars) can generate lower-bounded interpretations in some contexts. In contrast, there is no obvious

¹⁰⁴ I note that this explanation depends on number expressions being interpreted lower-boundedly when they are used as degree modifiers, as well as when they are used as quantifiers. That might not be automatically predicted by all of the accounts I discuss in §7.

linguistic support for FTT's claim that number expressions are represented as denoting *some*. Of course, FTT emerged as a psychological theory of memory, not of language; nevertheless, its application to risky-choice framing depends on a distinctively *linguistic* claim about how number expressions are interpreted. It is worth noting, then, the presence of linguistic counterevidence to the claim that number expressions are interpreted as being equivalent to 'some'. I will focus on a study by Panizza, Huang, Chierchia, and Snedeker (2015).

Using the 'visual world paradigm' devised by Huang and Snedeker (2009), Panizza et al. tasked their participants with matching spoken descriptions to items presented in a visual display. At the beginning of each trial, the (adult) participants were told a story by a teacher figure about four children (two boys and two girls), who were depicted in separate quadrants of the visual display. Participants were then asked by a cook figure, who could not hear this story, to help find a child to assist in the kitchen. The child for whom the cook was looking was described in terms of the quantity of an object s/he had, using either a quantifier ('some' or 'all') or a number expression ('two' or 'three'). For example, participants might be told: 'A boy has two of the paperclips. Point to him.' They would then respond by selecting the matching image from the visual display or, if there was no match, by clicking on a marked box to reject the sentence. Panizza et al. used a within-subjects experimental design; each participant heard eight sentences in each of the four conditions ('some', 'all', 'two' and 'three').¹⁰⁵

For current purposes, two types of trial were of critical importance: in the first type of trial, participants were asked to look for a boy (or girl) with 'some' of a given object, where one boy (or girl) had all three and the other had none. In this scenario, a lower-bounded reading of 'some' would enable a match with the child who had all three, whereas a bilateral reading would require the sentence to be rejected. In the second critical trial type, participants were looking for a boy (or girl) with 'two' of a given object but, again, one boy (or girl) had all three while the other had none. Again, a lower-bounded reading of 'two' would allow for a match with the child who had all three, whereas a bilateral reading would require the sentence to be rejected.

Panizza et al. found that, in the first critical trial type, using 'some', participants typically *did* match the description to the child who had all three items (74% did this in 'upward-entailing' grammatical contexts; 93% in 'downward-entailing' contexts). This suggests that the lower-

¹⁰⁵ Subjects were further split between conditions in which the scalar expression appeared in 'upward-entailing' or 'downward-entailing' grammatical contexts. 'Upward-entailing' contexts license inferences from sets to supersets; for example, '200 women were saved' entails '200 people were saved', where 'woman' is a subset of 'person'. In contrast, 'downward-entailing' contexts license inferences from sets to *subsets*; for example, 'It is not the case that 200 people were saved' entails 'It is not the case that 200 women were saved'. I will not discuss this aspect of the experiment any further here.

bounded reading, ‘some and possibly all’ was typically adopted, whereas a bilateral, ‘some but not all’ interpretation was either cancelled or never adopted in the first place.

In contrast, in the second critical trial type, using ‘two’, participants typically rejected the sentence (89% did this in ‘upward-entailing’ contexts; 60% in ‘downward-entailing’ contexts). This suggests that they preferred a bilateral, ‘exactly’, interpretation of the number expression, rather than maintaining or adopting a lower-bounded, ‘at least’, interpretation.

These results relate to the discussion of risky-choice framing in a couple of different ways. The point I wish to emphasise is that they are, at least *prima facie*, at odds with FTT. Presumably, FTT predicts that the responses in the ‘two’ trial should pattern similarly to those in the ‘some’ trial. This is because, when faced with a number expression as in the ‘two’ trial, FTT claims that people will reason with the ‘gist-like’ representation, ‘some’ (understood as ‘some and possibly all’), as long as that suffices for completing the task at hand. And in this case it does suffice: it is possible to complete the matching task by deploying ‘some and possibly all’, namely by matching the description to the child who has all three items. Importantly, even though participants *could* have reasoned with the ‘gist-like’ representation, clearly they did not. Instead, while participants typically made the match in the ‘some’ condition, they tended to reject the target sentence in the ‘two’ condition. This indicates that they were deploying a more precise interpretation of ‘two’.

A natural response for fuzzy trace theorists here would be to argue that some element of Panizza et al.’s experimental set-up prompted subjects to work with an unusually precise interpretation of ‘two’ rather than representing it as denoting *some*. This might be due, for example, to the experimenters having used *small* numbers. In a different study, Degen and Tanenhaus (2015) find that ‘some’ is considered unnatural when describing sets containing one to three items. Perhaps, then, ‘two’ is more likely to be represented verbatim than the larger numbers typically used in framing studies (like ‘200’ and ‘400’).¹⁰⁶ While this hypothesis has some *prima facie* plausibility, it would at least require a refinement of FTT, to restrict its claims to *sufficiently large* numbers. Pending further empirical investigation, however, there is little reason to suppose that even *large* numbers are preferentially represented as ‘some’.

Before moving on, I note that Panizza et al.’s experiment threatens to undermine the lower-bounding hypothesis: after all, if cardinal expressions in that study are typically given bilateral, ‘exactly’ interpretations, it might seem unwarranted to suppose that the number expressions

¹⁰⁶ Compounding this worry, the within-subjects design could perhaps have encouraged participants to look for a distinction between ‘two’ and ‘some’, exaggerating any pre-existing tendency to represent them differently.

used in framing studies generate lower-bounded readings. As a first point in response, even if only a *minority* of participants are reasoning with lower-bounded interpretations, this may be sufficient to contribute towards a framing effect, in line with the modest version of the lower-bounding hypothesis I endorse. Second, it is possible that scenarios like the ADP encourage lower-bounded readings to a greater extent than does Panizza et al.'s scenario (although this is clearly a claim that would require further empirical testing).

7. Semantics, Pragmatics, and Lower-Bounding

In the preceding discussion, I argued that framing effects probably depend in part on some participants interpreting number expressions lower-boundedly. However, this does not yet address the question of *why* they do so. In this final section I briefly survey various possible answers to that question, which trace lower-boundedness to the meaning (broadly construed) of an uttered number expression. On some views, lower-boundedness is seen as part of the *conventional* meaning of a number expression; on others it is held to be the result of a pragmatic effect, which enriches or modulates the conventional meaning.

According to the traditional neo-Gricean approach, developed in different ways by Horn (1989) and Levinson (2000), number expressions (and other scalars) conventionally have lower-bounded denotations. On this view, '200' in sentence (1) conventionally denotes *at least 200*, as in (1a), while '400' in sentence (2) conventionally denotes *at least 400*, as in (2a).

(1) If Program A is adopted, 200 people will be saved.

(1a) If Program A is adopted, *at least 200* people will be saved.

(2) If Program C is adopted, 400 people will die.

(2a) If Program C is adopted, *at least 400* people will die.

On this view, then, (1) and (2) are inequivalent with respect to their basic semantics, which could explain why they are represented differently.¹⁰⁷

¹⁰⁷ Of course, they need not be represented differently, since some participants might pragmatically enrich the conventional meaning, arriving at bilateral interpretations. A neo-Gricean explanation for such enrichment might go roughly as follows: if more than 200 people would be saved, the speaker would have said so explicitly (assuming the speaker to be cooperative and knowledgeable about the number of people who will be saved). However, since the speaker did not do that, it must be the case that no more than 200 people will be saved. Note that, since this conclusion relies on assumptions of cooperativity and knowledgeability, participants are not necessarily *required* to draw it: instead, they might prefer to leave open the possibility that the speaker does not know (or want to say) exactly how many people will be saved. This could potentially explain why some individuals adopt bilateral interpretations while others adopt lower-bounded ones.

A distinct semantic explanation of lower-bounded interpretations of number expressions takes number expressions to be ambiguous or polysemous. According to this kind of approach, number expressions have two lexical entries, or senses, as a matter of linguistic convention, one of which has a lower-bounded denotation, while the other has a bilateral denotation. To give an idea of how this works, I will briefly outline the polysemy-based analysis put forward by Geurts (2006).¹⁰⁸

Geurts argues that the basic conventional meanings of number expressions are bilateral senses. According to his semantic analysis, this means ‘200 people will be saved’ can be paraphrased roughly as follows:

(1d) There uniquely exists a set of exactly 200 people who will be saved.¹⁰⁹

According to (1d), the expression ‘200’ maps a predicate to *TRUE* just if that predicate holds of one and only one group of 200 people.¹¹⁰ In other words, the semantics of ‘200’ is bilateral and precise.

Geurts proceeds to show that when number expressions are used *predicatively* they require a different analysis. This is formally derived by applying to the basic meaning a standard type-shifting operation known as ‘Quantifier Lowering’ (Partee, 1987). Consider, for example, sentence (6) in which the number expression is used predicatively:

(6) These are 200 people.

By applying Quantifier Lowering to the basic sense of the number expression, it is possible to arrive at a semantic analysis of (6) that can be paraphrased as follows:

(6a) The number of these is exactly 200 and these are people.¹¹¹

In (6a) the expression ‘200’ is analysed as denoting a property – namely, a property of whichever sum of entities is denoted by ‘these’.¹¹²

Once this predicative sense of number expressions is added to our semantics, it becomes susceptible to a further standard type-shifting operation known as ‘Existential Closure’. Crucially, this operation maps number expressions to *lower-bounded* meanings, which can

¹⁰⁸ According to a related proposal, put forward by Chierchia, Fox, and Spector (2012), the different readings of number expressions arise from the interaction between grammatical context and a covert syntactic operator known as the ‘exhaustivity operator’. However, for reasons of space, I will not discuss this account.

¹⁰⁹ Formally: $\exists!x[\#x = 200 \wedge \text{People}(x) \wedge \text{Saved}(x)]$, where ‘ $\exists!x$ ’ means that there uniquely exists an x , and ‘ $\#x$ ’ means the number of entities in x .

¹¹⁰ Formally: $\lambda P \exists!x[\#x = 200 \wedge P(x)]$

¹¹¹ Formally: $\#these = 200 \wedge \text{People}(these)$.

¹¹² Formally: $\lambda P. \lambda x[\#x = 200 \wedge P(x)]$

therefore be posited as additional conventional senses. For example, applying Existential Closure to the predicative analysis of ‘200’ produces a semantic analysis of (1) that can be paraphrased as follows:

(1e) There exists a set of exactly 200 people who will be saved.¹¹³

Although the inclusion of ‘exactly’ in (1e) might appear to invoke a bilateral reading of the number expression, in fact it produces lower-bounded reading here. This is because the group of exactly 200 people is no longer required to be *unique*: ‘200’ in (1e) maps a predicate to *TRUE* just if the predicate holds of *at least one* group of exactly 200 people.¹¹⁴ Equivalently, it maps a predicate to *TRUE* just if it holds of at least 200 people.

Geurts claims that the bilateral sense of a number expression is the more usual, dominant, and preferred reading, at least when the number expression is used as a quantifier. Nevertheless, his account seems to allow that some hearers of frames could select the lower-bounded sense. Whether or not the selection of one sense is more *justified* than the other remains a further question; however, the onus falls on those who claim that framing effects are irrational to show that the lower-bounded sense should *not* be deployed in framing scenarios.

A third approach to the semantics of number expressions similarly denies that number expressions are straightforwardly and univocally bilateral. However, it is argued that their conventional meanings are altogether *underspecified*, neither having bilateral nor lower-bounded denotations. This view, which is naturally allied with the Contextualist approaches to meaning variation discussed in chapter 2, has been advocated by Kempson and Cormack (1981) and Carston (1988, 1998). It claims that, on each occasion on which a number expression is used, its underspecified conventional meaning must be pragmatically enriched in one way or another. Hearers will thus arrive at lower-bounded, bilateral, or upper-bounded interpretations. If the view is correct, it might predict that some participants in framing studies interpret the number expressions in frames bilaterally, while others interpret them lower-boundedly.¹¹⁵ As before, it would require much further justification to show that participants rationally ought to derive only *bilateral* interpretations.

A final approach, taken by Breheny (2008), considers number expressions to have bilateral conventional meanings. This account, which is diametrically opposed to the standard neo-

¹¹³ Formally: $\exists x[\#x = 200 \wedge \text{People}(x) \wedge \text{Saved}(x)]$.

¹¹⁴ Formally: $\lambda P \exists x[\#x = 200 \wedge P(x)]$

¹¹⁵ Recall that, on the Relevance Theoretic approach Carston endorses, this will depend on which interpretation happens to be the most *relevant*, i.e. which is the most readily accessible interpretation that is informative enough to justify the cognitive effort required to derive it. Since relevance will depend in part on each individual’s extant set of beliefs, it is consistent with this account that different participants may arrive at different interpretations.

Gricean approach, fits naturally with the assumption that framing researchers have standardly made about the meanings of the number expressions in frames. Even so, Breheny is fully cognisant of the fact that number expressions can obtain lower-bounded interpretations in certain contexts. Specifically, he appeals to the role of hearers' background knowledge in their derivation of lower- or upper-bounded readings. On a Breheny-style account, then, it might be hypothesised that differences in participants' background knowledge could affect their tendency to derive lower-bounded or bilateral readings of the number expressions in the frames they receive. Therefore, even this bilateral approach is, in principle, consistent with the ability of the lower-bounding hypothesis to provide a rationalising explanation of framing effects.

I will not arbitrate between the various positions on the meanings of number expressions, which have been briefly surveyed here. I do not believe the framing data bears directly on this large and complex debate.¹¹⁶ However, whether lower-bounded interpretations end up being traced back to conventional meaning or pragmatic enrichment, what I hope to have shown is that far more work would be required to show that hearers behave *irrationally* in deriving such interpretations. As things currently stand, then, the lower-bounding hypothesis stands to provide a partial explanation of framing effects, and one that may be compatible with hearers' responses being rational.¹¹⁷

8. Conclusion

I have argued that the lower-bounding hypothesis is likely to explain framing effects in part, though not in full. The data from Chick et al. (2016) show that, even when hearers represent number expressions bilaterally, a framing effect can still arise. In order to investigate the unexplained portion of the effect, I will now proceed as if the number expressions in alternative frames did have bilateral meanings, and were represented by participants as such. This will allow me to isolate the other factors contributing to the framing effect. With the assumption of bilaterality in place, then, I am now in a position to accept that many pairs of alternative frames are logically equivalent, at least at the level of their minimal semantic meanings. In other words, if we compose the conventional meanings of their constituent expressions, we arrive at meanings that share a single set of truth-conditions.

It is necessary to pause at this point because the assumption that frames, *qua* sentences, can even *possess* truth-conditions (let alone *share* them) immediately begs the question against Radical Contextualists, who prominently deny this. In the next chapter I address their objection.

¹¹⁶ For further discussion, see (Spector, 2013).

¹¹⁷ I will return to the issue of rationality in chapter 9.

CHAPTER 5

Truth-evaluability¹¹⁸

Abstract

The debate between Semantic Minimalism and Radical Contextualism is standardly characterised as concerning *truth-evaluability* – specifically, whether or not sentences require rich contextualization in order to express complete, truth-evaluable contents. In this chapter, I examine the notion of truth-evaluability, considering which kinds of mappings it might require from worldly states of affairs to truth-values. At one end of the spectrum, an exhaustive notion would require truth-evaluable contents to map *all* possible states of affairs to truth-values. At the other end, a liberal notion would require only that truth-evaluable contents map *at least one* possible state of affairs to *at least one* truth-value. I show that both Minimalists and Radical Contextualists rely on some intermediate, moderately strict notion of truth-evaluability, falling between these two poles. I consider four ways in which such a notion could be defined. However, I argue that each of these is ultimately implausible, giving us no reason to favour a moderately strict notion of truth-evaluability over the liberal alternative. This suggests that Minimalists and Radical Contextualists alike should accept that sentences can express truth-evaluable contents. I conclude that, in principle, the meanings of alternative frames can be understood as being truth-evaluable and capable of logical equivalence.

1. Introduction

The notion of truth-evaluability features centrally in the debate between Semantic Minimalism and Radical Contextualism. Recall from chapter 2 that Minimalists argue that well-formed sentences in language express complete, truth-evaluable contents, with only a very limited role for the contexts in which those sentences are uttered.¹¹⁹ Thus, Borg (2012, p. 3) writes:

According to minimal semantics, natural language sentences mean things, the things they mean are in some sense complete (that is to say, they are propositional, truth-evaluable contents), and these literal meanings are determined entirely as a function of the lexical elements a sentence contains together with its syntactic form.

¹¹⁸ This chapter draws heavily on my published paper, ‘Reassessing truth-evaluability in the Minimalism-Contextualism debate’ (Fisher, 2019).

¹¹⁹ Specifically, that role is exhausted by providing values for a limited set of genuinely context-sensitive elements, as will be discussed again in §3.

Cappelen and Lepore (2005) defend a similar view, arguing that sentences can express complete, truth-evaluable propositions, without pervasive contextual influences.¹²⁰

Radical Contextualists, on the other hand, deny that we can sensibly talk about *sentences* expressing complete, truth-evaluable contents. They argue instead that contexts must always play a substantial role, such that only richly contextualized *utterances* of sentences can express complete, truth-evaluable contents. Again, according to Travis (2008, p. 70):

The new view might be called the speaking-sensitive view of words, or language... [Words'] fixed, language-contributed semantics must, in general, be supplemented if they are to be properly assessable as to truth; that is, if they are to count either as true or as false. Nor is such supplementation provided automatically by some fixed stock of 'indexical features'.¹²¹

This in-principle claim, that minimal contents are necessarily incomplete, distinguishes Radical Contextualism from its more moderate neighbour. Moderate Contextualists accept that there can be minimal truth-evaluable contents but claim that these have no theoretical or practical importance in their own right. Instead, the contents of primary importance are held to be those which the speaker *intuitively communicates*; and these will often be highly context-dependent. As we saw in chapter 2, Recanati (2010) advocates such a view, under the label of 'Truth-Conditional Pragmatics'. He writes:

There is no doubt that one can *define* 'what is said' in such a way that only weak pragmatic effects can affect what is said. If we do so, Minimalism comes out true by definition...What the truth-conditional pragmatist means by 'what is said' corresponds to the intuitive *truth-conditional content* of the utterance. According to [Truth-Conditional Pragmatics], the intuitive truth-conditions of an utterance *are* affected by free pragmatic processes. (Recanati, 2010, p. 12, emphasis in original)¹²²

I will argue in §6 that a closer analysis of the notion of truth-evaluability makes it difficult to maintain the distinction between Radical Contextualism's in-principle claim and Moderate Contextualism's practical relevance claim. For now, though, let us return to the dispute between

¹²⁰ I will focus mainly on Borg's version of Minimalism here. However, in §3, I will recap some ways in which it differs from that of Cappelen and Lepore.

¹²¹ I will focus on the version of Radical Contextualism put forward by Travis (2006, 2008) although I note that Carston (2002) makes a similar claim.

¹²² More will be said in later sections to distinguish 'free' pragmatic processes from the 'weak' processes accommodated by Minimalists.

Minimalism and Radical Contextualism. Both sides have explicitly positioned truth-evaluability at the heart of their disagreement. As Borg (2012, p. 46) puts it:¹²³

Where minimalism differs from occasionalism then is over the question of whether [content recoverable without rich appeal to the context of utterance] can itself be thought of as genuine, truth-evaluable content. Travis...is adamant that it can't... It is this commitment to the idea that any content recovered simply on the basis of lexico-syntactic features alone (without any rich appeal to the context of utterance) is destined to fall short of propositional, truth-evaluable content that makes the occasionalist a true opponent of the minimalist.

Responding to Cappelen and Lepore, Travis conceptualises the debate similarly:

[Cappelen and Lepore] assign truth precisely that role in semantics which [Radical Contextualism] denies it. They insist that (unproblematic) declarative sentences—'Pigs grunt', 'There are French women in Chicago', 'The oboe is double-reeded'—and even more problematic ones—'Giraffes are tall', 'Sid has had enough'—express, *as such*, 'minimal propositions'. Whatever else a minimal proposition is, it is (for them) something truth-evaluable, and, as a rule, truth-valued. That, then, is the main issue between [Semantic Minimalism] and [Radical Contextualism]. (Travis, 2008, pp. 151-152, emphasis in original)

In this chapter, I do not seek to defend either a Minimalist or Contextualist stance. Instead, I will argue that it is a mistake to characterise the debate between Minimalism and Radical Contextualism in terms of truth-evaluability. I will show that, once this is acknowledged, the disagreement turns out to concern the *usefulness* of minimal truth-evaluable contents, rather than their existence per se. Accordingly, the meanings of frames *qua* sentences can be considered truth-evaluable and capable, in principle, of logical equivalence. In later chapters I will go on to argue that these minimal semantic meanings also have explanatory value.

The structure of the chapter is as follows: in the next section, I introduce a spectrum of possible notions of truth-evaluability. In §3 and §4, I show that both Minimalists and Radical Contextualists rely on *moderately strict* notions of truth-evaluability, falling somewhere between the liberal and exhaustive extremes. In §5, I consider—and reject—four possible candidates for a moderately strict notion of truth-evaluability. I conclude, in §6, by arguing that

¹²³ In this passage, Borg is using the term 'occasionalism' to describe what I am calling Travis's Radical Contextualism.

the debate should not be framed in terms of the truth-evaluability of contents but what various truth-evaluable contents *do* for us. This dispels the Radical Contextualist's in principle objection to the idea that sentences can have truth-evaluable meanings.

2. What Truth-evaluability Requires

I start from the assumption that what makes a content truth-evaluable is, in some sense, its ability to map worldly states of affairs to truth-values.^{124, 125} A naïve view would require truth-evaluable contents to map *all* possible states of affairs to truth-values. Take, for example, the following sentence:

(1) Serena Williams is American

According to the naïve view's exhaustive criterion, for sentence (1) to express a truth-evaluable content, states of affairs must either be ones in which Serena is American (and which are therefore mappable to *TRUE*) or ones in which that is not the case (mappable to *FALSE*). Crucially, no states of affairs may remain unmapped.

The naïve view, however, is typically considered to be implausibly stringent by philosophers on both sides of the debate. Borg (2012, p. 108) writes:

On this approach then a proposition *p* must suffice to determine, for every possible world, whether *p* is true or not in that world. As it stands, this condition is too strong (for instance, there are issues with worlds where objects fail to exist, worlds where fundamental laws are very different to those of this world, and perhaps with vagueness).

Travis (2008, p. 134) similarly points out: "Our language is not rigid. It does not foresee its own application in, or to, every circumstance in, or to, which it may prove to need to be applied".

Building on this, he writes:

¹²⁴ Therefore, should truth-evaluability be defined in an entirely different way, the arguments developed here might not apply.

¹²⁵ Putting things in terms of mappings effected by contents themselves allows us to target a metaphysical question, distinct from epistemic questions concerning language-users' *knowledge* or *beliefs* about those mappings. This strategy avoids begging the question against those who consider such epistemic questions to be independent of the truth-evaluability of contents themselves. Borg, for example, aims to keep semantic theory free of substantive epistemological concerns (Borg 2004, pp. 1–2). Such a strategy does not presume that contents can map states of affairs to truth-values *wholly* independently of a community of language-users; it rests only on the weaker point that such mappings may be independent of what *any given member of the community* happens to know or believe.

Take an arbitrary collection of statements, for each of which there is a way which is the way it said things to be. Is it guaranteed that every member of the collection is either true or false? Does that follow merely from those statements being the ones they are? Or from each having the feature that there is such a thing as how it said things to be? My answer is no. (Travis, 2008, pp. 142-143)

It seems clear, then, that neither Minimalists nor Radical Contextualists wish to deploy an exhaustive notion of truth-evaluability. Instead, they are thinking of truth-evaluable contents as achieving only some *partial* mapping of possible states of affairs to truth-values. The task is to say what this partial mapping should be.

One possibility lies at the opposite end of the spectrum. A liberal notion of truth-evaluability would count any content as truth-evaluable so long as it mapped *at least one* possible state of affairs to *at least one* truth-value. Returning to (1), there need only be some possible state of affairs in which Serena is American, or one in which that is *not* the case.¹²⁶

In §3 and §4, I will show that neither Minimalists nor Radical Contextualists can be thinking of truth-evaluability in this liberal way. Instead, both of their accounts depend on some stricter notion. Although there is relatively little discussion in the literature of how to define such a notion, in §5 I will consider four alternatives. These require truth-evaluable contents to map, respectively:

- (i) at least one possible state of affairs to *TRUE*;
- (ii) *many*, or *most*, possible states of affairs to truth-values;
- (iii) possible states of affairs to truth-values *in an intuitively correct way*;
- (iv) *actual* states of affairs to truth-values.

¹²⁶ I am grateful to Nat Hansen for pointing out that this liberal notion seems technically consistent with Bach's claim that a "complete and determinate proposition" must be "something capable of being true or false" (Bach 1994, p. 127). In practice, though, Bach seems to rely on a stricter notion. He deems many sentences to fall short of expressing complete, truth-evaluable propositions (thus expressing only 'propositional radicals'). These nevertheless seem capable of expressing something *false* (applying the strategy to be fleshed out in §§3 and 4 below). Although such sentences would satisfy the second disjunct of Bach's official criterion, he clearly considers them to be sub-propositional.

Jumping ahead for a moment, I will argue that none of these options is plausible. I will propose instead that we revert to the most liberal way of defining a mapping from states of affairs to truth-values.

3. Truth-evaluability in Minimalism

Minimalists appeal to truth-evaluability at two junctures in their account: first, to demonstrate that some sentences require (very limited) relativization to contexts, in order to express truth-evaluable minimal contents; and, second, to defend against any *further* contextual influences on such contents.

According to Minimalists, the truth-evaluable, minimal contents expressed by sentences are determined compositionally; roughly-speaking, by the meanings of words (as given by their entries in the lexicon) and the way they are combined (according to grammatical rules).

In typical cases, the contributions that sentence components make to minimal contents are held to be independent of the particular context in which they are uttered. Nonetheless, as discussed in chapter 2, Minimalists standardly recognize a set of exceptional elements in language that are genuinely context-sensitive. These include, at least, indexicals (like ‘I’ and ‘you’), demonstratives (like ‘this’ and ‘that’), a small set of other, open-class expressions (like ‘here’, ‘yesterday’, and ‘actual’) and tense-markers. Such context-sensitive elements can have different extensions in different contexts; for example, the expression ‘she’ can refer to different (female) individuals, when uttered by different speakers on different occasions.

Minimalists deny that sentences containing these kinds of context-sensitive elements can express truth-evaluable contents independently of context. Thus Borg writes:

To claim that pragmatics has no role to play at all within the semantic realm is very likely to lead to the claim that the subject matter of semantics is sub-propositional or non-truth evaluable content. The reason for this is pretty easy to see: consider the sentence ‘I’m here now’. Unless one is able to look to a specific context in which this sentence is uttered it will be impossible to deliver values for the context-sensitive terms ‘I’, ‘here’ and ‘now’. Yet without such values any content we can recover for the sentence (perhaps in terms of a Kaplanian character) is bound to fall short of propositional, truth-evaluable content. (Borg, 2012, p. x)

Take the following sentence:¹²⁷

(2) She is American

The context-independent, linguistic meaning of sentence (2) places some conditions on its truth. Specifically, it requires that there is at least one female individual who is American; we can express this as follows:¹²⁸

Context-Independent Content (CIC): ‘She is American’ is true **only if** $\exists x$ (x is female \wedge x is American).

Minimalists consider CIC to fall short of being a truth-evaluable content because it does not specify *which* individual—the actual referent of ‘she’ in the context of utterance of (2)—must have the properties of being female and American.

Relatedly, CIC is not a traditional truth-condition: it captures only *necessary* but not *sufficient* conditions for the truth of sentence (2) (the direction of the conditional travels only from left to right). This means that CIC is incapable of mapping any states of affairs to *TRUE*. Instead, for any state of affairs that includes at least one female American (thus satisfying the necessary condition expressed by CIC), the truth-value of the content remains indeterminate (neither definitely true nor definitely false). These considerations are taken as evidence against contents like CIC being truth-evaluable.¹²⁹

Note, though, that CIC is capable of mapping some states of affairs to *FALSE* – specifically, those in which there are *no* female Americans. The context-independent, linguistic constraints provided by sentence (2) are sufficient on their own to rule out such states of affairs. Recall too that the liberal notion of truth-evaluability described in §2 above requires only that *at least one* possible state of affairs be mappable to *at least one* truth-value. If we were to adopt this notion,

¹²⁷ In this sentence, ‘she’ is to be understood deictically rather than anaphorically. This ensures that its referent cannot be determined purely linguistically, within the discourse context, but only by appeal to features of the extra-linguistic context.

¹²⁸ According to some accounts (2) may express a truth as long as the speaker’s intended referent is American, even if that individual turns out not, in fact, to be female. Thus, the necessary condition in the consequent of CIC reduces to: $\exists x(x$ is American). If preferred, readers can simply plug in this alternative condition, and substitute ‘American’ for ‘female American’ throughout the ensuing discussion.

¹²⁹ Note that CIC differs from two other kinds of contents in the neighbourhood. First, it differs from a content expressible by the sentence, “There are some female Americans”, for which the condition ‘ $\exists x(x$ is female \wedge x is American)’ would be not only necessary but also *sufficient*. That content would map any state of affairs containing at least one female American to *TRUE*. CIC also differs from Perry’s ‘reflexive content’ (Perry, 2012). The reflexive content of (2) would typically determine a value for x ; the important point is that it would do so via a relation between x and the *utterance* of (2), not via the *direct reference* of ‘she’ to x .

then, CIC would count as a truth-evaluable content (just in virtue of its being ‘falsity-evaluable’, as it were). Since Minimalists deny that CIC is truth-evaluable, it is clear that they cannot be operating with the liberal notion of truth-evaluability. Instead, they are assuming something stricter. In §5, I will consider some potential candidates. For now, I continue setting out the positive Minimalist view, leading to its second appeal to truth-evaluability.

Minimalists argue that a sentence like (2) can only express a truth-evaluable content once relativized to a context of utterance. Specifically, that context must determine the referents of context-sensitive elements like ‘she’. To characterise the truth-conditions of minimal contents, Borg (2004, pp. 165-166; 2012, pp. 135-136) adopts Higginbotham’s conditionalized t-sentences (Higginbotham & Segal, 1994, pp. 92-93). For sentence (2), this gives us something like the following:

Minimal Content (MC): If the speaker of ‘She is American’ refers with the utterance of ‘she’ therein to *x* and to nothing else, then this sentence, as uttered in this context, is true **if and only if** *x* is American.

MC depends on a specific, contextualized utterance of (2), in order to incorporate *x* as the *actual* referent of ‘she’.¹³⁰ This is what supposedly allows MC to provide *sufficient* as well as *necessary* conditions on truth. Since MC is a standard biconditional truth-condition, it maps states of affairs to *TRUE* if *x* is American; and to *FALSE* if *x* is *not* American. In this way, Minimalists standardly acknowledge some role for context in determining minimal contents.

A crucial qualification is that the role for context is restricted to providing values only for *genuinely context-sensitive* elements of the sentence. As discussed in chapter 2, there is some disagreement between Minimalists as to how these genuinely context-sensitive elements should be defined. Cappelen and Lepore (2005) primarily seek to limit the *quantity* of context-sensitive elements in language (more or less to those enumerated by Kaplan (1989)). Borg (2004, 2012) instead prioritises a *qualitative* constraint, requiring context-sensitivity to be traceable back to lexico-syntactic features of the sentence. The key idea here is that any context-sensitivity must already be built into the lexical entry or grammatical rule.¹³¹

¹³⁰ Specifically, ‘she’ is supposed to acquire the value, *x*, via *direct reference*. This means that MC2 is not a reflexive content in Perry’s sense (Borg, 2012, p. 136, footnote 27).

¹³¹ In this way, it is maintained that minimal contents flow directly from a sentence’s formal, compositional, semantic meaning. As we will see in §5, though, they need not correspond to *communicated* contents, which may also depend on pragmatic factors. As Kratzer (2012, p. 4) puts it:
Words, phrases, and sentences acquire content when we utter them on particular occasions. What that content is may differ from one context to the next. It is the task of semantics to

Setting this wrinkle aside, Minimalists typically agree that a minimal content like MC is truth-evaluable as it stands, without further contextual enrichment. The claim that there are complete, truth-evaluable, minimal contents of this kind constitutes Minimalism's second appeal to truth-evaluability. As will be discussed in the next section, it is also the focus of the Radical Contextualist critique.

Before moving on, though, it should be noted that Minimalists may well have other reasons for accommodating limited contextual influences on minimal content, aside from considerations of truth-evaluability. In particular, they point out that certain context-sensitive elements of language are *referring* expressions, and rigidly designate their referents (see, for example (Borg, 2012, pp. 134-142)). Whereas a content like MC captures this feature of the expression 'she', CIC does not; in that sense, CIC might be deemed 'incomplete'.

It would take us too far afield to evaluate the substance of claims about which expressions are rigid designators. Instead, I want only to show that such claims are orthogonal to the current discussion. The question we are concerned with here is whether or not a content like CIC is *truth-evaluable*, not whether it fails to capture certain facts about the meaning of the expression 'she'. Minimalists have claimed that one problem with a content like CIC is that it falls short of truth-evaluability; it is *this* claim I wish to interrogate. I allow that a different incompleteness challenge may be mounted against CIC, on the basis that it fails to respect certain meaning facts. Indeed, this is an approach I discuss further in §§5 and 6. However, once the Minimalist takes this path, any appeal to truth-evaluability, as such, simply falls away.¹³²

4. Truth-evaluability in Radical Contextualism

Radical Contextualists claim that minimal contents, and not just the likes of CIC, inevitably fall short of truth-evaluability. Extending the Minimalist's argument against CIC, it is argued that the context-independent, linguistic meanings of sentences *always* underspecify truth-evaluable contents.¹³³ Travis writes:

describe all those features of the meaning of a linguistic expression that stay invariable in whatever context the expression may be used. This invariable element is the meaning proper of an expression.

¹³² In §5 I will also reject attempts to incorporate such meaning facts within the very definition of truth-evaluability itself.

¹³³ This is not a straightforward *generalization* of the Minimalist's argument, since Radical Contextualists think the incompleteness arises in a different way: whereas, for Minimalists, context-sensitivity can only emerge from lexical or syntactic features, Contextualists argue for *free* contextual enrichment, unbidden by linguistic features. I return to discuss this distinction again in §6.

The driving force of [Radical Contextualism] is this idea: the open sentences of language speak of ways for things to be which admit of understandings... This blocks truth-conditional semantics. For suppose I say, ‘The sentence “Sid grunts” is true iff Sid grunts’. Either I use that last ‘grunts’ on some particular understanding of being a grunter—one understanding among many—or I do not. If I do, then I assign the sentence a property it does not have. For *it* does not speak of being a grunter on any special understanding of this. But if I do not, then I fail to state *any* condition under which anything might be true. Being a grunter on no particular understanding of being one is just not a way for Sid to be. In brief, the choices here are falsehood or failure to say anything. What would be needed to block this result are ways for things to be, which one might speak of, and which do not admit of understandings. (Travis, 2008, p. 159, emphasis in original)

Similar reasoning can be applied to the expression ‘American’ in our sentence (2). Imagine, for example, a tennis player who is an American citizen but has nonetheless entered a tournament as part of a team that is unaffiliated to America. In a context in which the player’s citizenship is at stake, an utterance of (2) might be considered true. However, in a context in which her team affiliation is at stake, an utterance of the same sentence might be considered false. According to the Radical Contextualist, there are *no* available understandings of expressions like ‘grunt’ and ‘American’, in the absence of specific contexts of utterance.^{134,135}

From the Radical Contextualist’s perspective, then, the minimal contents posited by Minimalists fall short of truth-evaluability, since they fail to recognise the full extent of contextual effects. With respect to sentence (2), the minimal content can be thought of as placing only a *necessary* condition on the truth of an utterance of (2). Specifically, it requires that the referent of ‘she’ is American on *some* understanding of ‘American’—an understanding which could potentially be

¹³⁴ In response, Minimalists typically reject this claim outright. They argue, for example, that such expressions pick out properties common to all individuals of whom they may be literally and truthfully predicated. See, for example, (Cappelen & Lepore, 2005, chapter 11) and (Borg, 2012). For opposing views, see (Chomsky, 2000; MacFarlane, 2007; Pietroski, 2005, 2018).

¹³⁵ Since the issue here concerns the satisfaction of predicates like ‘grunts’ and ‘is American’, it would be possible to reframe the discussion in terms of the satisfaction-conditions of such predicates, rather than the truth-conditions of sentences that contain them (an anonymous reviewer of (Fisher, 2019) raises this point). In order to address the Minimalist and Contextualist arguments head-on, I have opted to retain the standard framing in terms of truth-evaluability. However, I do not believe anything of substance hangs on this decision; the same considerations that favour a liberal notion of *truth-evaluability* (enumerated in §5) would also favour a liberal notion of *satisfiability*, such that a predicate would count as satisfiable so long as at least one thing were ruled into, or out of, its extension.

made available in some context or other. The truth-condition must therefore be pared back to the following conditional:

Minimal Content 2 (MC2): If the speaker of ‘She is American’ refers with the utterance of ‘she’ therein to *x* and to nothing else, then this sentence, as uttered in this context, is true **only if** there is some understanding of ‘American’ on which *x* is American.

Whereas the Minimalist’s MC assumes that there is some context-independent understanding of ‘American’, the Radical Contextualist denies that any such understanding is available. In the absence of an appropriate, contextually-determined understanding of ‘American’, MC2 fails to provide a value for that expression. As before, this leaves us with a content that lacks a sufficiency condition for its truth. It is incapable of determining whether *x* is American, and therefore cannot map any states of affairs to *TRUE*. Instead, states of affairs in which *x* is American on some possible understanding remain unmapped to either truth-value. This is taken to show that MC2 falls short of truth-evaluability.

Nevertheless, again, MC2 is capable of mapping some states of affairs to *FALSE* – specifically, those in which *x* is not American, on *any* possible understanding. This seems compatible with Travis’s account; in the long passage quoted below, he argues that linguistic meaning places important constraints on expressible content:

I think that the English ‘is round’ speaks, as such, of being round. So I think it speaks of that on every use of it which is (a case of) speaking proper English. I am not dissuaded from this view by the fact that, on different such speakings of it, it will make different contributions to the truth-conditions of wholes of which it may, then, be part. There is a fact about that bit of English which I hope I can capture in that way. It is that the meanings of those words constrain, in a particular way, what you can say on an occasion in using them—even if they do not narrow things down to just one thing. I can, occasion permitting, call the squash ball round and speak truth of it even as it begins its rebound off the wall. But I cannot so easily call it round and speak truth of it just in case it barks, or is made of lead, or is on fire. My idea is: I can use ‘is round’, on an occasion, of a ball, to say what I would say on that occasion in *calling it round*; and (*ceteris paribus*, perhaps) that is all I can use those words to say of it. Further, to say what I just did is to say how the meanings of those words constrain their use. I cannot use those words of the ball (speaking English, so that they mean what they do) and thereby say the ball to be on fire, unless there is an understanding of being round on which to be round is to be on fire. (Which, so far as I can see, there is not.) I do think this is a pretty stringent constraint on

what you can use ‘is round’ to say in speaking English. (Travis, 2008, pp. 175-176, emphasis in original)

I interpret the linguistic constraints described by Travis as ruling out *some* possible states of affairs.¹³⁶ As we have seen, this is all that is needed for contents to count as truth-evaluable on a liberal notion of truth-evaluability. Radical Contextualists, then (like Minimalists) must be operating with some suitably stricter notion of truth-evaluability, given their denial that contents like MC2 are truth-evaluable.

5. Stricter Notions of Truth-evaluability

In the previous two sections, I demonstrated that neither Minimalists nor Radical Contextualists can be working with a liberal notion of truth-evaluability. Instead, both camps are relying on something stricter. There has been little explicit discussion of what that stricter notion is. In this section I will consider four possibilities that require truth-evaluable contents to map, respectively:

- (i) at least one possible state of affairs to *TRUE*;
- (ii) *many*, or *most*, possible states of affairs to truth-values;
- (iii) possible states of affairs to truth-values *in an intuitively correct way*;
- (iv) *actual* states of affairs to truth-values.

I will argue that none of these works. Of course, given the form of the argument, this does not prove that no moderately strict notion of truth-evaluability is available, since more promising candidates might yet be identified. However, by eliminating some initially plausible possibilities, I aim to shift the burden of proof to those who rely on such a notion, to provide a workable definition.

Starting with (i), we saw that contents like CIC and MC2 were incapable of mapping any possible states of affairs to *TRUE* (even though they map some to *FALSE*). The first suggestion, then, would require truth-evaluable contents to map at least one possible state of affairs to *TRUE*, not

¹³⁶ Likewise, if we chose to think about things in terms of predicate satisfaction, a predicate’s invariant satisfaction-condition would act as a constraint, or filter, preventing at least some things from being included within its extension on any occasion of use (and therefore ensuring that at least some possible applications of the predicate would express falsehoods). This is despite the fact that the satisfaction condition might not define any very specific extension that the predicate may be used to pick out on a particular occasion of utterance.

just to *FALSE*. In other words, this proposal stipulates that the truth-condition must be a standard biconditional, preventing mere ‘falsity-evaluability’ from entailing truth-evaluability.

The problem with this proposal is that it would automatically deny truth-evaluability to *necessary falsehoods*. By definition, these cannot map any states of affairs to *TRUE* (no state of affairs could correspond to the right-hand side of a biconditional truth-condition). Take, for example, a sentence of the following form:

(3) x is F and not-F.

Let’s assume that the lexical entry for ‘F’ has no context-sensitivity built into it; and that both occurrences of ‘F’ in (3) obtain identical understandings in the context of utterance. In such a scenario (3) will express a contradiction, mapping no states of affairs to *TRUE*. Nevertheless, I take it that both Minimalists and Radical Contextualists would consider it to express something truth-evaluable (just false).¹³⁷

It might be objected, however, that necessary falsehoods are relevantly different from the contents expressed by sentences like (1) and (2). An obvious difference is that necessary falsehoods map *all* possible states of affairs to *FALSE*, leaving no possible worlds unmapped. It might be suspected, then, that the problem with the liberal notion of truth-evaluability is that it leaves some possible worlds unmapped to either truth-value.

As discussed in §2, neither Minimalists nor Radical Contextualists endorse an exhaustive notion of truth-evaluability, so the problem can’t be a simple failure to map *all* possible states of affairs to truth-values. Perhaps, though, contents like CIC and MC2 map *too few* states of affairs to truth-values. This brings us to proposal (ii), which would require truth-evaluable contents to map *many*, or *most*, possible states of affairs to truth-values. However, any such criterion is at best vague and at worst unworkable. It is vague as long as the requirement to map ‘many’ states of affairs to truth-values fails to specify *how many*. Any attempt to impose such a threshold would seem straightforwardly arbitrary. A deeper problem is that, assuming there are infinitely

¹³⁷ Note that I am not requiring Radical Contextualists to sign up to the idea that sentence (3) expresses something necessarily false independently of context. I have explicitly allowed for the possibility that it does so only following rich contextualization, as long as both occurrences of ‘F’ obtain identical understandings in the context. It is important here to distinguish between the *context of utterance*, which determines the content of (3), and the *context of evaluation*, which determines whether or not (3) is satisfied. (3) may express a necessary falsehood only in some contexts of utterance, but wherever it does so, it will be false in all contexts of evaluation. This point comes up again below, in relation to option (iv).

many possible states of affairs, it may not even make sense to talk about some *proportion* of them being mappable to truth-values.

Another way of interpreting the worry is that the contents expressible with sentences like (1) and (2) (unlike necessary falsehoods) should, intuitively, map certain kinds of states of affairs to *TRUE* – specifically, those in which the individual being referred to is, in fact, American (whether ‘American’ is thought of as expressing a property context-independently, or only on a context-specific understanding). This brings us to suggestion (iii). The idea is that truth-evaluable contents must map states of affairs to truth-values in an *intuitively correct* way, i.e. in a way that reflects what a speaker could reasonably mean by using those words.

I take it that CIC and MC2 would not count as truth-evaluable on this basis. A speaker uttering (2) is reasonably taken to refer to *some particular (female) individual* by using ‘she’; and to predicate *some property* of that individual by saying that she is ‘American’. Thus, intuitively, the content of (2) should map to *TRUE* those states of affairs in which the individual in question has the relevant property; and it should map to *FALSE* those in which she doesn’t.^{138,139}

Nevertheless, intuitions about appropriate mappings to truth-values would need to be of the right kind for this proposal to get going. For the Minimalist, that would mean showing that the intuitions concern the *minimal* content of the sentence, rather than some contextually richer *speech act* content that gets communicated by the sentence in use. In general, Minimalists are careful to point out that minimal contents need not (and typically won’t) answer to all of our intuitions about what is communicated by an utterance. As Borg (2012, p. 48) puts it:

These minimal contents provide the literal meaning of sentences relativized to contexts of utterance, though, the minimalist acknowledges, they do not usually provide the intuitive contents of speech acts involving those sentences.¹⁴⁰

To see the point, take the following sentence, as uttered by a tennis spectator to her companion:

¹³⁸ I note that the question of how widely these intuitions are shared is ultimately an empirical matter.

¹³⁹ This may be one way of cashing out a suggestion we find in Borg, that one might consider a content truth-evaluable “just in case it makes a ruling in a range of clear-cut possible scenarios” (Borg, 2004, p. 238). Note that Borg does not endorse this suggestion; she argues only that *if* it were adopted, minimal contents would count as truth-evaluable.

¹⁴⁰ Similar points are made by Cappelen and Lepore (2005) and García-Carpintero (2013). It is worth noting that each of these philosophers is careful to emphasize the very important role played by speech act contents in a theory of linguistic communication. Their claim is therefore not that we can, or should, disregard such contents. Rather, it is that they should be held apart from minimal contents, which play a distinct theoretical role.

(4) Williams bounced the ball three times and served an ace.

What the spectator intuitively communicates is that Williams bounced the ball three times and *then* served an ace. However, the minimal content of the sentence is consistent with Williams having performed those actions in the reverse order, i.e. serving an ace and then bouncing the ball three times. Minimalists thus place a wedge between a sentence's minimal content (based on compositional semantics) and the wider (pragmatic) speech act contents of utterances.

At most, then, minimal contents aim to capture only some *subset* of intuitions about what speakers could reasonably mean; namely, those which reflect the constraints imposed by context-independent, linguistic meanings. These would need to be carefully delineated before the Minimalist could appeal to what speakers of (2) intuitively use 'she' to mean. Specifically, it would need to be shown that the use of 'she' to refer to a particular female individual relates to the expression's *lexical entry* rather than to use-specific facts.¹⁴¹

Assuming that the relevant intuitions can be pinned to a sufficiently restricted content, still they seem to arise from meaning facts that are quite distinct from the requirements of truth-evaluability. For example, it is because 'she' is taken to be a certain kind of referring expression that the content of (2) is expected to concern the referent of 'she' (and to be true if and only if *that particular individual* is American). Rather than stipulating that truth-evaluable contents must respect such facts, it seems more parsimonious to appeal to the facts directly when arguing for contextual relativization. As discussed in §3, then, the Minimalist might argue that contents like CIC are incomplete, on the basis that the lexical entries for expressions like 'she' demand contextually-determined referents. Yet this is entirely consistent with CIC being *truth-evaluable*. The argument that CIC fails to capture our intuitions about the invariant meaning of the expression 'she' is independent of that content's status as truth-evaluable or not. In sum, nothing is gained by building independently motivated intuitions into the notion of truth-evaluability.¹⁴²

¹⁴¹ Radical Contextualists may face a similar problem if they wish to distinguish the content directly expressed by the utterance from other contents that are merely indirectly expressed, or implicated. Specifically, it would need to be shown that the use of 'American' to pick out a context-specific property relates to what the speaker of (2) directly expresses, rather than indirectly implicates. There is some evidence that Travis sees value in maintaining such a distinction; see, for example, his discussion in 'On What is Strictly Speaking True' (Travis, 2008, chapter 1).

¹⁴² Analogously, from the Contextualist perspective, nothing is gained by conflating a content's truth-evaluability with whether a speaker intuitively communicated it. Thus, even if there is a sense in which speakers are reasonably taken to mean something fairly specific by using the word 'American', we can appeal to that fact without saying that considerations of *truth-evaluability* require 'American' to have such a specific extension.

Finally, option (iv) would require any truth-evaluable content to map actual states of affairs to truth-values. Again, this would preclude contents like CIC and MC2 from being truth-evaluable. For example, given that the actual world does contain female Americans, CIC fails to map any actual states of affairs to truth-values (not even to *FALSE*).

This condition is also implausible, however. At the same time as denying the truth-evaluability of the sentence ‘She is American’, it would deem the structurally similar sentence ‘She is demoniac’ to be perfectly truth-evaluable, in virtue of its being mappable to *FALSE* in the actual world (assuming there are in fact no demons). This seems to carve up the territory in the wrong way, for both the Minimalist and the Radical Contextualist.^{143,144}

To sum up, I currently see no reason to favour a moderately strict notion of truth-evaluability over the liberal alternative. The liberal notion has the advantage of being modest without being vacuous: it successfully captures the idea that truth-evaluability concerns mappings of states of affairs to truth-values, without imposing additional, unjustified, constraints on the nature of those mappings. Pending the identification of more promising candidates for a moderately strict notion, I conclude that we should take contents like CIC and MC2 to be truth-evaluable, in line with the liberal notion.

6. Implications

Where does this leave us? First, acknowledging that contents like CIC are truth-evaluable undermines the Minimalist’s claim that sentences like (2) require relativization to contexts, in order to be truth-evaluable. (As we will see below, however, those sentences may still be argued to require contextual relativization on separate grounds.)

Likewise, acknowledging that contents like CIC and MC2 are truth-evaluable neutralizes the Radical Contextualist’s claim that sentences can only ever express truth-evaluable contents once richly contextualized. (Again, though, as will be discussed below, it may still be denied that such contents have any explanatory value.)

¹⁴³ The underlying problem seems to be that option (iv) makes a content’s *truth-evaluability*—not just its truth-value—depend on how the actual world happens to be. Instead, *truth-evaluability* is better thought of as a function of content, which (whether richly dependent on the context of utterance or not) must be fixed independently of its satisfaction in actual or possible worlds. Thus, only the truth-value of the content may change across different contexts of evaluation, not its *truth-evaluability* in general.

¹⁴⁴ Perhaps a fifth option could require truth-evaluable contents to map, not *actual* states of affairs, but some *contextually-relevant* set of possible states of affairs, to truth-values. Indeed, something along these lines is suggested by Rayo (2013). However, getting the proposal off the ground would require, at least, some non-arbitrary way to establish which states of affairs are in the contextually relevant set.

A third point, which falls out from the first two, is that the debate between Minimalism and Radical Contextualism cannot fundamentally concern *truth-evaluability*. I have suggested that both camps should agree that minimal contents are truth-evaluable. As I will argue below, I think the debate between them ultimately concerns the *practical significance* of minimal contents.

In fact, Minimalists and Radical Contextualists shouldn't only agree that minimal contents are truth-evaluable but also that contents like CIC are truth-evaluable. In other words, there is a level of truth-evaluable content that is even more minimal than Minimalism's minimal content. I will call this 'marginal content'.¹⁴⁵

Although I think Minimalists should recognize that marginal contents like CIC are truth-evaluable, I note that this is consistent with their maintaining that minimal contents like MC retain an important role in semantic theorising. Such contents, recall, capture just those contextual effects which are anticipated by an expression's lexical entry, or a sentence's combinatory syntax. Thus, I believe the key task facing the proponents of standard versions of Minimalism, is to demonstrate the explanatory value of the minimal contents they posit, which hold apart 'weak' or 'mandatory' contextual effects from the 'free' contextual effects that influence wider speech act contents. Minimalists have indeed sought to meet this challenge; see, for example, (Borg, 2004, 2012, 2017) and (Cappelen & Lepore, 2005). Although I do not attempt to assess their arguments here, they seem to me to be the cornerstone of any defence of the view.

On the other side, I have argued that Radical Contextualists should also recognize both minimal and marginal contents as truth-evaluable. In other words, they should drop the in-principle claim that sentences cannot express truth-evaluable contents independently of context. That means Radical Contextualism must take up a more moderate position, objecting to minimal contents purely on the basis that that they lack *practical relevance*, since they generally fail to capture what speakers communicate.

¹⁴⁵ The proposal here is in some ways similar to that of García-Carpintero (1998), which identifies minimal semantic content with context-independent, linguistic meaning. However, García-Carpintero assumes that this content is not truth-evaluable. Others who restrict 'semantics' to an entirely context-independent level of meaning similarly take it to fall short of truth-evaluability (in some or all cases) – see, for example (Carston, 2002) and (Harris, forthcoming). My purpose here has been to show that semantic constraints of the kind envisaged are, in fact, sufficient for truth-evaluability. Of course, though, we can (and should) posit additional levels of meaning in order to capture other aspects of what is *communicated* by uttered sentences.

Indeed, this line of argument is commonly pursued by Contextualists of all stripes. As we saw in §1, Moderate Contextualists acknowledge that sentences express truth-evaluable, minimal contents, while arguing that these contents have no significance, in their own right. Thus, whereas appeals to truth-evaluability have standardly purported to distinguish Radical Contextualism from Moderate Contextualism, they in fact end up obscuring a key point of agreement between the two views.

A key task facing both Moderate and Radical Contextualists, then, is to argue (against the Minimalist) that mandatory contextual effects have no privileged status in a theory of meaning but, instead, the contents of theoretical interest are those which can be *freely* influenced by the context. For some arguments to this effect, see (Carston, 2002) and (Recanati, 2004, 2010). Again, I cannot assess these here but I suggest that they are central to the dispute between Minimalists and Contextualists. It is only possible to recognize this point once we dispel the spectre of truth-evaluability. The upshot is that we should focus on what the contents posited by competing theories *do* for us, not whether they are *truth-evaluable*.

As an aside, my claim that Radical Contextualism collapses into a more moderate position may seem to contradict arguments put forward by Cappelen and Lepore (2005) and Borg (2012) to the effect that it is *Moderate* Contextualism that is doomed to collapse into *Radical* Contextualism.¹⁴⁶ I believe that the disagreement here is, in large part, only apparent. In this chapter, I have focused narrowly on the different parties' claims about *truth-evaluability*; it is with respect to *these* claims that I have argued Radical Contextualism must move towards a more moderate stance. In contrast, Cappelen and Lepore (2005, chapters 3-5) focus on Contextualists' claims about the number of context-sensitive expressions in language. They argue that, once the Moderate Contextualist begins to expand the set of context-sensitive expressions, there is no stable place to stop, until all linguistic expressions are included. Thus, taking the first step on a Contextualist path inevitably sends one down the slippery slope to a far more radical position. Whether or not they are right about this, it remains entirely consistent with what I have said; sliding towards the more extreme position on their scale, concerning the pervasiveness of context-sensitivity, does not imply taking a radical stance on the question of whether minimal contents can, in principle, be truth-evaluable. As I have attempted to show, a sentence can still express something truth-evaluable, despite its containing context-sensitive elements, just so long as its invariant meaning filters out *some* possible states of affairs.

¹⁴⁶ I am indebted to an anonymous reviewer of (Fisher, 2019) for helping me see this.

Meanwhile, Borg (2012, pp. 35-38) addresses Contextualists' claims about 'sharpening' or 'precisifying' contents, in order to capture what a speaker communicated in context. Whereas Moderate Contextualists standardly take such sharpenings to be possible, some Radical Contextualists have argued (and Borg agrees) that the process is, in principle, never-ending: the only way to stop making increasingly fine-grained distinctions is by acknowledging that context plays an *ineliminable* role in content-determination. She thus identifies a different way in which Moderate Contextualism seems inexorably to slide toward a more radical position. Again, though, this would be consistent with my argument since, even if one is forced to accept that only richly contextually-embedded contents capture what a speaker communicated, this would not imply that such contextualization is required for *truth-evaluability*. Thus, for example, according to Travis's Occasion-sensitive account, it is impossible to sharpen communicated contents in a way that would allow them to be represented, either in language or in thought (this distinguishes his approach from many other forms of Contextualism). However, this dimension of his view is quite independent of the claim that only contextually-embedded communicated contents can be *truth-evaluable*. Again, it is the latter claim that has been my target here.

As mentioned in chapter 2, the question of whether it is possible to represent communicated contents accurately in language, abstracting from context, is one which falls beyond the scope of the thesis. Throughout my discussion of framing effects, I provide linguistic formulations that are intended to capture roughly how uttered frames might be interpreted. If Travis is right, these formulations will never fully capture what is actually communicated. Be that as it may, they are sufficient for my purpose, which is to investigate why alternative frames are interpreted *differently* from one another. Therefore, I will not address the radical representational claim that is distinctive of Travis's Occasion-sensitivity. I will only address (in chapter 8) a claim it shares with other, more moderate, forms of Contextualism – namely, that minimal semantic meanings lack explanatory value.

7. Conclusion

At the end of the previous chapter, I raised a possible objection to my claim that some pairs of alternative frames are logically equivalent, at least at the level of their minimal semantic meanings. The Radical Contextualist would deny that frames, *qua* sentences, can have meanings that are even truth-evaluable, let alone logically equivalent to one another. In this chapter, I have argued that this is mistaken, based on a closer examination of the notion of truth-evaluability. I have argued for a liberal notion of truth-evaluability, which would block the Radical Contextualist's in-principle objection to minimal contents. In conclusion, then, I maintain that alternative frames *can* have logically equivalent semantics. In the next two chapters, I will develop the idea that they nevertheless convey different information at a *pragmatic* level.

CHAPTER 6

Information Leakage

Abstract

In this chapter, I endorse the ‘information leakage’ account of attribute framing effects, according to which alternative frames carry distinct information about the utterance context. I consider whether that information might concern reference points, implicit recommendations, psychological salience, or argumentative orientation. I also explore a related account, which focuses on the information conveyed by hearers’ judgements, rather than by speakers’ frames. I provide empirical and theoretical considerations in favour of the *reference point hypothesis*, which is developed and refined for deployment in the next steps of my argument.

1. Introduction

From this point on, I shift my focus from risky-choice framing to attribute framing. As discussed in chapter 1, attribute framing is a simpler paradigm, in which different wording is used to attribute the same property to an entity. Take for example, the following pair of sentences, used to describe a basketball player:

(1) This player made 40% of his shots last season.

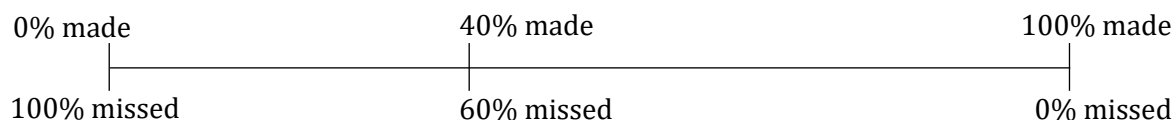
(2) This player missed 60% of his shots last season.

As argued in chapter 3, ‘made’ and ‘missed’ have contradictory meanings here: since no shot can be both missed and made, the expressions denote mutually exclusive sets; and, since any shot is either missed or made (with no third possibility) the sets they denote are also jointly exhaustive. In line with the discussion in chapter 4, the proportions, 40% and 60%, will be assumed to have bilateral, complementary meanings. Taken together, these properties of frames (1) and (2) ensure that they have logically equivalent meanings: if one composes the conventional meanings of their constituents, the resulting meanings will be true (or false, or indeterminate) in the same possible worlds. The conclusion of chapter 5 was that we need not deny that frames *qua* sentences have such meanings, so long as truth-evaluability is understood sufficiently liberally. Therefore, attribute frames like (1) and (2) can indeed be understood as having logically equivalent minimal semantic meanings.¹⁴⁷ I therefore define attribute frames as follows:

¹⁴⁷ And, indeed, logically equivalent ‘marginal’ meanings, in the terminology of the previous chapter. For simplicity, I will stick to discussing their ‘minimal’ meanings, in which context-sensitive elements are assumed to be saturated.

Attribute frames: pairs of sentences that predicate the same property of an individual or object by using contradictory predicate expressions together with complementary quantificational expressions.

According to their minimal semantics, then, both (1) and (2) locate the basketball player's performance at exactly the same point on the scale below.



Intuitively, though, the player sounds better when described using (1) rather than (2). As mentioned in previous chapters, experimental results confirm this intuition: a player is deemed more valuable to a team when described using (1) rather than (2) (Leong et al., 2017). In this chapter, I adopt the following definition of the attribute framing effect:

Attribute framing effect: a systematic shift in audience judgements, brought about by the speaker's choice of attribute frame (while all aspects of the wider context are held constant).

The structure of the chapter is as follows: in §2, I outline the dominant 'associationist' account of attribute framing effects. In §3 I argue that this account cannot explain framing effects which do not depend on the evaluative polarity of alternative frames. In §4 I describe a competing approach, known as the 'information leakage' account, which can capture evaluatively-neutral framing paradigms. In §5 I explore a specific version of the information leakage account, known as the 'reference point hypothesis'. In §§6-8 I assess this hypothesis against alternatives based on 'implicit recommendations', 'psychological salience', and 'argumentative orientation'. In §9 I consider a related account developed by Geurts (2013), which focuses on the information conveyed by hearers' judgements rather than by speakers' frames. I conclude by endorsing a refined version of the reference point hypothesis, which I take to be currently the most theoretically and empirically well-supported version of the information leakage account.

2. The Associationist Account

Recall from chapter 1 that an expression will be considered to have positive or negative evaluative valence insofar as it denotes something that is typically thought of as good or bad, respectively. Assuming that it is good for a basketball player to *make* shots (given that this is the aim of the game), frame (1) can be thought of as positively worded, in virtue of its inclusion of 'made'. Conversely, frame (2) can be thought of as negatively worded, since it is a bad thing for a

basketball player to *miss* shots. As noted in chapter 1, I will restrict my focus to *affirmative* uses of valenced expressions like ‘make’ and ‘miss’. I thus leave aside for now the question of whether negating these verbs would produce a frame that is positive, negative, or neutral overall.

Data from attribute framing research confirms that positively-worded attribute frames lead to relatively *favourable* evaluations of the entity being described, whereas their negatively-worded counterparts lead to relatively *unfavourable* evaluations (for a survey of the experimental literature, see (Levin et al., 1998)). Noting this ‘valence-consistent shift’, Levin and others have proposed that a speaker’s wording being positive or negative directly affects how hearers encode the information conveyed (Levin, 1987; Levin & Gaeth, 1988; Levin et al., 1986; Levin, Johnson, Russo, & Deldin, 1985; Levin et al., 1998). They write:

Attribute framing is likely to influence the encoding and representation of information in associative memory, and this representational difference is viewed as the cause of *valence-consistent* shifts in responses. (Levin et al., 1998, p. 164, emphasis in original)

To see roughly how this is supposed to work, when a hearer of (1) encodes the semantically-determined information that the basketball player made 40% and missed 60% of his shots, this is encoded with some kind of positive association(s). Levin et al. do not say much about the mechanics of the encoding process; perhaps we could think of the information as receiving a positive ‘tag’, or as being stored in memory within a network of positively-evaluated information.

Either way, when the information about the player’s performance is subsequently retrieved from memory, it comes with the relevant positive association(s). This is held to colour the hearer’s judgement in the evaluative task, explaining why the player is judged more valuable to a team in the positively worded framing condition than in the negatively worded condition.

Again, it remains rather unclear exactly how the positive association(s) are supposed to affect the judgement. In places, the authors seem to suggest that attention plays a mediating role, such that the positive association(s) would cause hearers of (1) to attend mainly to the positive attributes of the player, rather than the negative ones (Levin et al., 1998, p. 167). I find this proposal rather puzzling, though: since alternative frames predicate the *same* attribute of an entity, there are no distinct (positive vs. negative) attributes to attend to under each frame.

For now, I will simply grant that there could be some mechanism through which positive associations attaching to information retrieved from memory lead to relatively favourable evaluations of the entity that information concerns (and, conversely, negative associations

encoded under negatively-worded frames can lead to relatively *unfavourable* judgements of the entity).

Since the positive or negative associations evoked by attribute frames are supposed to bias hearers' subsequent judgements, the associationist account conceives of attribute framing effects as a species of *priming* effect. What makes attribute framing effects particularly distinctive – and particularly robust – is how explicit the priming is. As Levin et al. put it:

In attribute framing, the framed stimulus label is like a prime, but it is part of, rather than peripheral to, the description of the target stimulus. It is no wonder that the effect is so reliable. (Levin et al., 1998, p. 166)

I will not attempt a detailed assessment of the associationist account here; evidently, several important aspects remain in need of further elaboration. Instead, I will argue that it fails to explain a subset of framing effects (whereas these *can* be captured by the competing 'information leakage' account, to be introduced in §4). In particular, as I show in the next section, framing effects can arise independently of frames being positively- or negatively-valenced.

3. Evaluative Neutrality

I will focus on an experiment reported by McKenzie and Nelson (2003). In this experiment participants were found to make systematically different judgements about the prior contents of a glass, based on how its current contents were described. Below is an example vignette, which was presented to participants in one of six experimental conditions:

Imagine that Mary was sitting at her kitchen table with a glass in front of her. She left the room briefly and came back to find that the contents of the glass had changed. When asked to describe the glass now, Mary said, "The glass is half full." Given how Mary chose to describe the glass after its contents had changed, please choose the statement below in terms of what you think was most likely true about the glass before its contents changed.

Participants were then given the following options:

The glass was full before its contents changed.

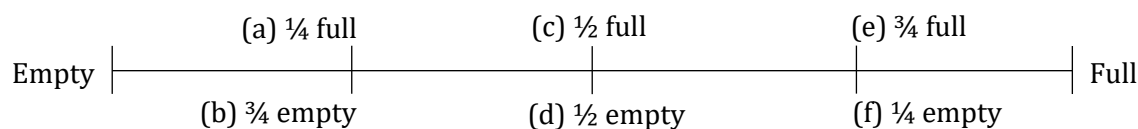
The glass was empty before its contents changed.

In two of the other conditions, Mary used the degree modifiers 'a quarter' or 'three quarters' rather than 'half'. These three 'full' framing conditions were mirrored by conditions in which

Mary used the contrary adjective, 'empty', with the complementary degree modifier. All six frames are listed below:

- (a) The glass is a quarter full.
- (b) The glass is three quarters empty.
- (c) The glass is half full.
- (d) The glass is half empty.
- (e) The glass is three quarters full.
- (f) The glass is a quarter empty.

The semantic information expressed by each frame is represented on the scale below.¹⁴⁸



There are three pairs of frames: (a) and (b); (c) and (d); (e) and (f). Within each pair, both frames describe the same *current* contents of the glass. Nevertheless, participants were found to make systematically different judgements about the *prior* contents of the glass, depending on whether Mary used a 'full' or 'empty' frame. Participants in conditions (a) and (c) (where Mary used 'full' frames) were more likely to say that the glass was previously empty than participants in conditions (b) and (d) (where Mary used 'empty' frames). And, of course, it follows that the reverse is true: participants in conditions (b) and (d) were more likely to say the glass was previously full than participants in conditions (a) and (c). The difference between conditions (e) and (f) fell just short of statistical significance, although the trend was in the same direction. The full results are presented in the table below:

¹⁴⁸ According to a prominent theory of gradable adjectives, 'full' and 'empty' are *absolute* gradable adjectives with 'closed' scales, i.e. scales with fixed endpoints (Kennedy & McNally, 2005). Therefore, a glass's being 'full' simpliciter means that it contains water to a maximal degree, while its being 'empty' simpliciter means that water is absent from the glass to a maximal degree. Once 'full' and 'empty' are combined with degree modifiers, however, they enable mappings to points along a scale, similarly to 'made' and 'missed' in the earlier example.

| Frame | % participants saying the glass was <i>full</i> before | % participants saying the glass was <i>empty</i> before |
|-------------------------|--|---|
| (a) $\frac{1}{4}$ full | 35% | 65% |
| (b) $\frac{3}{4}$ empty | 94% | 6% |
| (c) $\frac{1}{2}$ full | 50% | 50% |
| (d) $\frac{1}{2}$ empty | 80% | 20% |
| (e) $\frac{3}{4}$ full | 56% | 44% |
| (f) $\frac{1}{4}$ empty | 79% | 21% |

Notably, this framing study does not depend on the evaluative polarity of the frames. On the one hand, it must be acknowledged that describing a glass as ‘half full’ has acquired idiomatic status in English, connoting a positive outlook on a state of affairs (whereas describing a glass as ‘half empty’ connotes a negative outlook). On the other hand, the experimental task is independent of these evaluative connotations: participants are not asked to provide an *evaluative* judgement but rather to assess the probable prior contents of the glass. Thus, even if information about a glass described as part ‘full’ were encoded with some positive association(s), as per the associationist account this would not imply that the glass was more likely to have started out empty rather than full. Nor would information about a glass described as part ‘empty’ being encoded with some negative association(s) imply that it was more likely to have started out *full*.¹⁴⁹ Nevertheless, the experimenters observed a systematic shift in audience judgements, brought about by the speaker’s choice of ‘full’ or ‘empty’ frames. In other words, they found an attribute framing effect, without this being a ‘valence-consistent shift’.

McKenzie and Nelson’s findings concerning ‘full’ and ‘empty’ frames have been corroborated by subsequent experiments conducted by Sher and McKenzie (2006) and Ingram, Hand, and Moxey

¹⁴⁹ Other experiments use pairs of frames that are more straightforwardly evaluatively-neutral. For example, Honda and Yamagishi (2017) ask participants to judge a protagonist’s initial beliefs about the number of red balls in a box of coloured balls, based on his/ her later description of the probability of drawing a red ball. Below is an example vignette:

In the box, there are 100 red and white balls that have been thoroughly mixed. You are going to draw a ball from this box. You cannot check the contents of the box. One of your friends has checked the contents of the box, and found that there were 50 red balls. Then, the friend said, “Because there are 50 red balls in this box, it is likely that you will draw a red ball”.

The participants were then asked:

Which do you think the friend thought – that the number of red balls was more than 50 or fewer than 50?

In this case, drawing a red ball does not denote a positive event; nor does describing the outcome as ‘likely’ connote something good. Nevertheless, framing effects were observed for alternative descriptions of the probability of drawing a red ball (such as ‘likely’ vs. ‘quite doubtful’), despite these descriptions being mapped to similar numerical probabilities (for example, ‘likely’ and ‘quite doubtful’ were assigned mean probabilities of 49% and 47%, respectively).

(2014) (both of which will be discussed in more detail in chapter 7). These framing effects cannot be captured by the associationist account, since they do not plausibly depend on hearers having encoded the semantic information expressed by frames with positive or negative associations. In the next section, then, I turn to a competing account which can provide a unifying explanation of both valenced and evaluatively neutral framing effects.

4. Information Leakage

Over the last 20 years, McKenzie, Sher, and others have developed an explanation of attribute framing effects which they dub the ‘information leakage’ account (Leong et al., 2017; McKenzie & Sher, 2008; Sher & McKenzie, 2006, 2011). While they acknowledge that alternative frames have logically equivalent *explicit* meanings, it is argued that each frame implicitly conveys divergent information. As McKenzie puts it:

[R]ecent evidence has shown that a speaker’s choice among logically equivalent attribute frames can implicitly convey (or “leak”) normatively relevant information about the speaker’s reference point, among other things. (McKenzie, 2004, p. 874)¹⁵⁰

In the next two chapters I will argue that the distinction between explicitly-conveyed, logically equivalent information, on the one hand, and implicitly-conveyed, inequivalent information, on the other, is naturally and fruitfully cast in terms of the distinction between semantics and pragmatics. In chapter 9, I will return to the idea that the information frames ‘leak’ is *normatively* relevant. For now, though, I want to explore the basic proposal that alternative frames convey distinct information. This is importantly different from the associationist hypothesis that the frames convey the *same* information, which merely gets encoded with positive or negative associations. In the remainder of the chapter, I consider what additional information a frame might convey, beginning with the hypothesis that it concerns a ‘reference point’.

5. The Reference Point Hypothesis

According to the reference point hypothesis, by including one of two contradictory predicate expressions, an attribute frame tends to convey that the subject of predication possesses the corresponding property *to a relatively high degree*. In the proponents’ own words:

McKenzie and Nelson (2003) hypothesized the following regularity in linguistic behavior: (1) In describing a fixed state of proportionate affairs, speakers are more likely to describe the proportion in terms of “X1” when X1 has *increased* relative to the

¹⁵⁰ I interpret the central ‘leakage’ metaphor as an unofficial disclosure of information (whether deliberate or inadvertent), rather than as a *loss* of information.

reference point proportion (the norm, or what one would have expected) than when X1 has *decreased* relative to the reference point. (2) Listeners are sensitive to this regularity—that is, listeners are capable of correctly inferring the reference point proportion from the speaker’s choice of proportion-frame. (Sher & McKenzie, 2006, p. 471, emphasis in original)¹⁵¹

To see how this is supposed to work, consider again the following pair of frames from McKenzie and Nelson’s experiment:

(a) The glass is a quarter full.

(b) The glass is three quarters empty.

The reference point hypothesis predicts that (a) will tend to convey the ‘reference point information’ in (ai):

(ai) The glass is *relatively full*.

Conversely, (b) will tend to convey the reference point information (bi):

(bi) The glass is *relatively empty*.¹⁵²

Since the experimental vignette and task instructions make the *prior state* of the glass salient, that can be taken as the reference point on this occasion, relative to which the glass is now relatively full or relatively empty. As we will see, in other cases different kinds of reference points may be salient – for example, the reference point might correspond to an average, norm, or expectation. I will not seek to explain the determinants of salience here.¹⁵³ However, in the relatively simple experimental scenarios I discuss, the reference point is usually fairly obvious.¹⁵⁴

¹⁵¹ Their formulation of the hypothesis here implies a temporal component, since the proportion is described as *having increased* relative to a reference point. However, in other studies where they appeal to the reference point hypothesis, there is no temporal relation between the reference point and the subject of predication. For example, in the study by Leong et al. (2017) the reference point is assumed to be an average, or typical, individual. Therefore, I will understand the hypothesis in more temporally neutral terms.

¹⁵² Strictly speaking, the reference point information here should be the contents of beliefs attributed to the speaker, as in: *The speaker believes that* the glass is relatively full; and *The speaker believes that* the glass is relatively empty. For current purposes, I will assume that the hearer makes a justified inference from these more complex metarepresentational forms to the simpler ones provided in (ai) and (bi) in the main text. However, I will return to interrogate this assumption in chapter 9.

¹⁵³ Plausibly, salience could depend on a plethora of factors, including the wider discourse context, the nature of the subject of predication, and also subjective factors of the kind discussed by Verheyen, Dewil, and Égré (2018).

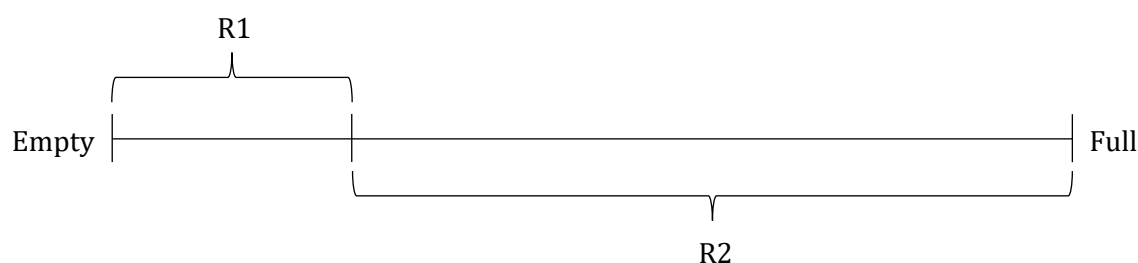
¹⁵⁴ More complicated cases are discussed in chapter 9.

Given the salience of the prior state of the glass in McKenzie and Nelson’s scenario, we can predict the following, more specific, reference point information to be conveyed by (a) and (b):

(aii) The glass is *fuller than before*.

(bii) The glass is *emptier than before*.

If (a) and (b) conveyed the information in (aii) and (bii), that could explain why, when Mary utters (a), participants are more likely to think the glass started out empty than when she utters (b) (and vice-versa). The idea is that alternative frames tend to establish diametrically opposed relations between the subject of predication and the reference point: in this example, frame (a) conveys that the prior state of the glass lies in region R1 on the scale below, while frame (b) conveys that it lies in region R2.¹⁵⁵



Depending on exactly what counts as instantiating a property to a ‘relatively high’ degree – i.e. *how much more* of the property an entity must instantiate, relative to the reference point – the regions R1 and R2 could turn out to be smaller than those depicted here. However, I will make the simplifying assumption that, so long as the entity being described possesses the relevant property to *some* higher degree than the reference point, then it counts as possessing it to a ‘relatively high’ degree.¹⁵⁶

Unlike the associationist account considered earlier, the reference point hypothesis can explain evaluatively neutral framing effects like the one observed by McKenzie and Nelson. As Sher and McKenzie write:

¹⁵⁵ It is not entirely clear why the experimenters used maximal rather than relative states in the options presented to participants, with the glass initially being *completely* full or *completely* empty, rather than simply *fuller* or *emptier*. Nevertheless, their results are still consistent with the reference point hypothesis.

¹⁵⁶ We might ultimately need to appeal to something like the notion of ‘standing out’, discussed by Kennedy (2007), such that the distinction is based on a *noticeable* difference in degree. For the sake of simplicity, I will set this issue aside for now.

[A]n associative account makes no predictions about the reference point phenomena..., which deal with valence-neutral descriptions and non-evaluative behaviors. (Sher & McKenzie, 2006, p. 482)

Additionally, the reference point hypothesis is easily extended to the valenced cases that dominate the framing literature. Returning to consider the example of the basketball player, I will now describe in more detail the first experiment by Leong et al. (2017). Below is an example vignette presented to participants in one of the ‘frame’ conditions:¹⁵⁷

Imagine that you are a recruiter for a college basketball team. Your job is to search for promising high school basketball players and try to recruit them to your college. You are looking through files for players from local high schools, and you are especially interested in players who can score many points.

The file you are currently looking at shows a player whose performance is quite unusual. This player made 40% of his shots last season.

Participants in the other ‘frame’ condition were told that the player ‘missed 60%’ of his shots. All participants were then asked:

How valuable do you think this player would be to your basketball team?

In line with previous framing research, the experimenters found a statistically significant valence-consistent shift, such that the player was judged to be more valuable under the positive ‘made’ frame than under the negative ‘missed’ frame. To see how this is explained by the reference point hypothesis, note that this account predicts (1) to convey (1a):

(1) This player made 40% of his shots last season.

(1a) This player made *a relatively large proportion* of his shots last season.

Conversely, (2) is predicted to convey (2a):

(2) This player missed 60% of his shots last season.

(2a) This player missed *a relatively large proportion* of his shots last season.

Under frame (1), then, the player is understood to have performed relatively *well* (assuming, of course, that making more shots is better). In contrast, under frame (2) he is understood to have performed relatively badly. Therefore, he is evaluated more favourably under (1) than under (2). Note that while both frames are taken to convey the same information about the player

¹⁵⁷ I will discuss the other ‘modeler’ and ‘recipient’ conditions below.

himself, they shift our assumptions about the reference point's location on the scale, so that the player's performance ends up seeming better or worse *by comparison*. Stated generally, the same subject can seem better under one frame than another just in virtue of being evaluated relative to different alternatives.

Leong et al. (2017) provide further support for the claim that it is reference point information which drives the attribute framing effect here. They investigate the extent to which the adoption of reference point information is *necessary* and *sufficient* for the effect to occur. Their first experiment addresses the sufficiency claim. In addition to the 'frame' conditions discussed above, this experiment included 'recipient' conditions, in which participants were told that a particular basketball player made 40% of his shots *and* missed 60% last season. In other words, they received the performance information in both frames. They were also given information about the performance of a 'typical' high school basketball player (again, this was presented in both 'made' and 'missed' frames). For some participants, the typical player's performance was better than that of the target player; for others the typical player's performance was *worse*.¹⁵⁸ The results from the recipient conditions showed that the better the typical player's performance, the less valuable the target player was judged to be. In other words, providing reference point information *explicitly* was sufficient to shift evaluative judgements.

Perhaps more surprisingly, the results of the second experiment suggest that framing effects are weakened where reference point information conflicts with a hearer's extant knowledge. Participants with varying amounts of prior basketball knowledge were asked to assess the value of a high school basketball player to a National Basketball Association (NBA) team. The player was either described as having 'made 60%' or 'missed 40%' of his free-throws; this represents fairly poor performance in an NBA context. It was found that participants who knew more about basketball were less susceptible to framing effects. In other words, the strength of the framing effect varied inversely with participants' basketball knowledge. The same participants were nevertheless equally susceptible to framing effects under a different, medical scenario. The experimenters take this as evidence that framing effects causally depend, at least partly, on the uptake of reference point information.¹⁵⁹

¹⁵⁸ The specific performance statistics for the 'typical' players were taken from estimates made by a different group of 'modeler' participants, who had judged average performance based on receiving information about the *target* player in one or other frame. However, for current purposes what is most important is that a range of higher and lower averages were provided.

¹⁵⁹ Since the effects were not eliminated entirely among knowledgeable participants, uptake of reference point information may not be *necessary* for framing effects to arise. Leong et al. appeal also to the 'implicit recommendations' that the frames convey (to be discussed in §6 below). In subsequent chapters, I will also return to consider in more depth the implications of reference point information being *defeasible*.

A series of other experimental studies, deploying various different pairs of frames, provide further support for the reference point hypothesis (Honda & Yamagishi, 2017; McKenzie & Nelson, 2003; Sher & McKenzie, 2006; Teigen & Karevold, 2005). Each of these studies also demonstrates that *speakers* choose their frames on the basis of what they believe about contextual reference points. In other words, it is not just that hearers infer reference point information but that, in doing so, they are tracking speakers' beliefs. I will discuss some of the data concerning speaker behaviour in the next chapter. In the remaining sections of this chapter, I want to consider some competing hypotheses within the broad 'information leakage' camp.

Before moving on, though, I end this section by providing the following refined characterisation of the reference point hypothesis:

Reference Point Hypothesis: In virtue of their including one of two contradictory predicate expressions, attribute frames tend to convey that the subject of predication instantiates the property corresponding to that expression to a high degree, relative to the most salient reference point.

6. Implicit Recommendations

Alongside the reference point hypothesis, Sher and McKenzie (2006) put forward an 'implicit recommendation hypothesis', according to which attribute frames carry information about the speaker's attitude towards the subject of predication. They write:

The fact that a speaker has described an object in terms of a positively valenced property makes it more likely that the speaker has a favorable attitude towards the object on the dimension under discussion. (Sher & McKenzie, 2006, p. 482)

The idea is that the positive or negative wording tends to reflect the speaker's overall, positive or negative, attitude towards a target state of affairs. So, for example, the choice to use (1) would indicate that the speaker has a favourable attitude towards the player, conveying to the hearer something like the information in (1b).¹⁶⁰

(1) The player made 60% of his shots.

(1b) The player is good.

¹⁶⁰ Again, strictly speaking, (1b) should be '*The speaker believes that the player is good*'. Again, I will assume for now that the hearer makes a justified inference from this metarepresentational form to the simpler one provided in the main text.

and groundbreaking attempts at seemingly unsolvable problems” (Sher & McKenzie, 2006, p. 484) as compared with the “incoherent attempts at simple problems” that constitute most of the poor team’s failures (Sher & McKenzie, 2006, p. 485). Conversely, the poor team’s successes are described as “slightly confused but nonetheless adequate” (Sher & McKenzie, 2006, p. 485), in contrast with the good team’s “truly extraordinary, sometimes revolutionary” ones (Sher & McKenzie, 2006, p. 484). Thus, the successes of the good team count more clearly as genuine successes, rather than mere avoidances of failure (which the poor team’s nominal ‘successes’ appear to be). Conversely, the failures of the poor team are more clearly failures (and don’t just fall somewhat short of being successes, as do the ‘failures’ of the good team). It is perhaps unsurprising, then, that these are the frames typically chosen by participants: the nature of the ‘successes’ and ‘failures’ themselves makes one frame clearly more fitting than the other. The experimental design does not allow us to exclude the possibility that this is what drives participants’ different responses.

Arguably, stronger support for the implicit recommendation hypothesis can be found in the wider literature. For example, several studies find that participants’ judgements about speakers’ attitudes systematically track the speakers’ uses of positively- or negatively-valenced frames (Karevold & Teigen, 2010; Teigen & Brun, 1999; Teigen & Nikolaisen, 2009). However, again, competing interpretations of the data are available. For example, participants could be projecting their own attitudes onto the speaker, rather than inferring the speaker’s attitude directly from the frame used. A similar point is made by Karevold and Teigen (2010, p. 725): finding a correlation between participants’ own attitudes and those they attribute to speakers, the researchers note the standard difficulties in inferring causation in one or other direction.

Importantly, the data also remain consistent with the reference point hypothesis: it is possible that hearers are first picking up on reference point information and then using that to infer the speaker’s attitude, as was suggested in the case of the basketball player. This possibility points to a conceptual problem facing the implicit recommendation hypothesis, which is explored next.

6.2. Evaluative Neutrality Revisited

Recall that a key motivation for the information leakage account, over the associationist alternative, was its ability to explain framing effects in evaluatively neutral cases (like the experiments involving the glass) as well as valenced cases. As discussed in §5, the reference point hypothesis meets this desideratum. In contrast, the implicit recommendation hypothesis only applies to framing effects that depend on the frames’ evaluative polarity. To that extent, it is unclear why the implicit recommendation hypothesis should be preferred to the associationist account.

The information leakage theorists may be thinking of the reference point hypothesis and the implicit recommendation hypothesis as forming a ‘package deal’; taken together, they are supposed to provide a better account than the associationist alternative. I want to suggest, though, that the implicit recommendation hypothesis may be superfluous. To reach that conclusion, it is helpful to unpick exactly how the information leakage theorists conceptualise their two hypotheses.

First, it is worth noting that there is some confusion about how the hypotheses are supposed to relate to one another. In places, the proponents suggest that the implicit recommendation hypothesis is *more general* than the reference point hypothesis. For example, Leong et al. write:

Related work has generalized the information leakage framework from signaling a speaker’s reference point...to signaling a speaker’s attitude toward the object—a type of implicit recommendation (Leong et al., 2017, p. 1155).

Perhaps, then, the idea behind the implicit recommendation hypothesis is that positive frames convey that target entities are *generally better* than the most salient comparator, whereas negative frames tend to convey that they are *generally worse*. The target entity would now be considered relatively good in potentially *any* respect, not just in virtue of having more of the good property corresponding to the predicate expression used. For example, rather than (1) tending to convey that the player *made more shots* than average, it would be taken to convey that the player *is generally better* than average.

Again, though, this generalising approach would restrict the scope of the information leakage account to cases where alternative frames are positively or negatively valenced, failing to capture evaluatively neutral framing effects. A more charitable view, then, would be that the two hypotheses are relatively *independent*, each capturing some portion of the framing data. Accordingly, Sher and McKenzie sometimes talk about frames being able to convey two distinct kinds of contextual information, as in the following quote:

[E]xperimental work has identified two...background conditions – reference points and implicit recommendations – that have broad relevance in attribute framing tasks. (Sher & McKenzie, 2011, p. 44)

In other words, framing effects may depend on reference point information *or* evaluative attitude information (or perhaps a combination of both).¹⁶² This would seem to be the best

¹⁶² Note that, on this understanding, only the reference point hypothesis need involve a comparative element. Under that hypothesis, frames convey information about the *relative quantities* of certain properties instantiated by target and reference point entities. The implicit recommendation hypothesis however, may concern purely *qualitative* information that frames convey about an entity’s goodness or

interpretation of the theorists' considered view. Additionally, though, they argue that both hypotheses can be subsumed under another, more general, one. I turn to that proposal in the next section.

7. Psychological Salience

Sher and McKenzie (2006) propose that, in general, frames convey information about whatever is 'psychologically salient' to the speaker. In cases where it is salient to the speaker that an entity instantiates a particular property to a relatively high degree, the frame will convey reference point information (as per the reference point hypothesis).¹⁶³ In other cases, the frame could be conveying various other sorts of information. In general terms:

(1) A speaker is more likely to describe *D* in terms of 'X1' when X1 is salient in the speaker's psycholinguistic representation of *D*.

(2) Listeners are sensitive to this regularity—they are more likely to (implicitly or explicitly) infer that X1 is salient in the speaker's representation of *D* when the speaker describes *D* in terms of 'X1'.

(McKenzie & Sher, 2008, p. 87)

Although this proposal has intuitive appeal, it is rather difficult to operationalise. After all, there are indefinitely many possible reasons why a given predicate expression could become salient to a speaker. Precisely which information a frame conveys on any given occasion, then, would be extremely unconstrained. Therefore, in the absence of a fuller account of psychological salience, the hypothesis has limited power to predict or explain framing effects. The theorists themselves recognise this, with Sher and McKenzie (2006, p. 486) noting that "the notion of psychological salience is a rough one, requiring further specification".

For the time being, I believe the reference point hypothesis remains the more theoretically well-developed and empirically well-supported version of the information leakage account.

Moreover, since it could explain framing effects in both valenced and neutral cases, it renders the implicit recommendation hypothesis potentially redundant.¹⁶⁴ That said, if the psychological salience approach could be worked out, I believe the arguments presented in the next three

badness *simpliciter*. In other words, the implicit recommendation hypothesis need not predict that an entity described using a positive frame is *better* than some comparator but merely that it is *good*.

¹⁶³ Sher and McKenzie (2006, p. 483) claim that "unusually abundant properties and events are more salient than unusually absent ones", although they do not offer independent empirical support for this.

¹⁶⁴ How, then, should we explain the residual framing effect found by Leong et al. (2017), even among hearers who were familiar with the typical performance of NBA basketball players? I believe various possible explanations are available here, aside from appealing to implicit recommendations. For example, hearers might assume that the speaker has some other reference point in mind. I return to this idea in chapter 9.

chapters of the thesis could be adapted accordingly, and would still stand. Before concluding the current discussion, I will consider two further proposals put forward by theorists working outside the core information leakage camp. The first of these concerns the ‘argumentative orientations’ of frames.

8. Argumentative Orientation

Holleman and Pander Maat (2009) propose that framing invokes a conversational principle of ‘argumentative orientation’ (a concept they inherit from Ducrot). The idea is that a frame, by including (or, in their terminology, ‘profiling’) one of two contradictory predicates, licenses a hearer to draw certain conclusions, based on the mere presence of the corresponding property (not its relative *abundance*, as under the reference point hypothesis). This is summarised in following pair of maxims:

Speaker’s maxim: when a situation lends itself to description in terms of a two-valued variable, profile the component carrying the value that best fits the direction of the conclusions one would prefer to be drawn from the utterance.

Recipient’s corollary: when a situation lends itself to description in terms of a two-valued variable, the component that is profiled indicates the direction of the conclusions the speaker would prefer to be drawn from the utterance.

(Holleman & Pander Maat, 2009, p. 2209)¹⁶⁵

In the formulation of these maxims, it is rather unclear what is meant by the ‘direction’ of conclusions one would prefer to be drawn. Indeed, the relationship between alternative frames and the distinct inferences they license is precisely what stands in need of explanation. One way of fleshing out the account would be to appeal to something like *representativeness*, whereby frames tend to convey that the subject of predication is representative of subjects that instantiate the relevant property. For example, describing a basketball player in terms of the number of shots he *made* might convey that he is generally representative of players who make their shots (being talented, tall, hard-working, or whatever). Conversely, describing him in terms of the number of shots *missed* might convey that he is generally representative of players who *miss* their shots. I believe this would be an interesting possibility to investigate further but, for reasons of space, it is not something I can pursue here. As things stand, though, the argumentative orientation account remains underspecified.

¹⁶⁵ Holleman and Pander Maat also posit a moderating effect of markedness. I leave this issue aside for now, although I will come back to it in the next chapter.

Setting aside this conceptual point for a moment, I will briefly run through Holleman and Pander Maat's empirical investigation of frames' argumentative orientations. Across six experiments, participants were presented with a series of sentence continuation tasks, most of which were language *production* tasks, as in the example below:¹⁶⁶

Choose the sentence that best fits next.

Next week I'll have much (little) time to come over and help you with odd jobs about the house.

(A) My diary is half full

(B) My diary is half empty

The experimenters predicted that having 'much' (or 'a lot of') time would be more likely than having 'little' time to prompt response B (and that having 'little' time would be more likely to prompt response A).¹⁶⁷ This prediction was borne out by the results. Indeed, responses patterned in the expected way for 30 out of a total of 34 production tasks.¹⁶⁸

Some participants also received *interpretation* tasks, in which they were asked to choose continuations on the basis of a speaker's choice of frame. An example is provided below:

Choose the sentence that best fits next.

My diary is half full (half empty).

(A) So I will have quite a lot of time to come do odd jobs about your house.

(B) So I will have little time to come do odd jobs about your house.

The experimenters predicted that, when the diary is described as 'half full', hearers would be more likely to infer that they have 'little' time than if it is described as 'half empty' (and vice-versa). The expected pattern was observed in all four interpretation tasks.

Although the results of the study are presented in support of the argumentative orientation hypothesis, I believe they remain consistent with the reference point hypothesis. For example,

¹⁶⁶ In the English translation the phrasing sounds slightly awkward. The stimuli were originally presented to participants in Dutch.

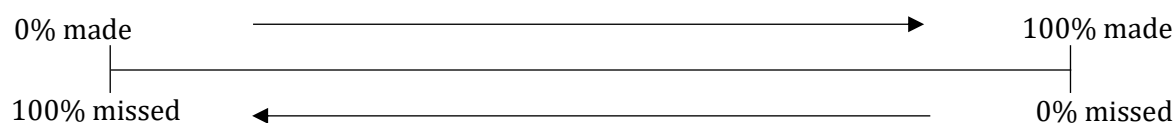
¹⁶⁷ If we were thinking about things in terms of representativeness, the idea here might be that describing the diary as 'half full' conveys that it is generally representative of a *full* diary, whereas describing it as 'half empty' conveys that it is generally representative of an *empty* diary.

¹⁶⁸ The expected pattern subsequently emerged for one of the remaining four tasks, when it was presented to a larger number of participants. Regarding the other three tasks, the authors speculate that choices may have been driven by other aspects of the particular scenarios used, concerning the expected end state, or goal.

inferring that one's diary is *relatively full* (under the 'half full' frame) could justify the conclusion that one will have relatively little time. And, conversely, inferring that it is *relatively empty* (under the 'half empty' frame) could justify the conclusion that one will have a relatively large amount of time. Therefore, I do not believe there are currently good empirical reasons to prefer the argumentative orientation hypothesis over the conceptually clearer reference point hypothesis.

9. Hearers' Responses

In this section, I consider one further account, which focuses on what *hearers* convey when they respond to speakers' uttered frames. Geurts (2013) begins by observing that alternative frames reverse the order in which related possibilities are located along a scale. For example, describing a basketball player in terms of the proportion of shots he *made* orders the following scale from left to right, with 0% of shots being made at the leftmost extreme, and 100% being made at the rightmost extreme.



Conversely, describing the player in terms of the proportion of shots he *missed* reverses the scale: at the rightmost extreme, 0% of shots are missed; at the leftmost extreme, 100% are missed.

Geurts proposes that frames evoke alternative formulations, corresponding to different points along the scale, i.e. 'scalar alternatives'. Such alternatives are standardly appealed to by linguists in explaining various communicative effects. As Geurts notes:

The interpretation of an utterance is determined not only by what the speaker says, but also...what he could have said (Geurts, 2013, p. 5)¹⁶⁹

The set of scalar alternatives to (1), for example, would include (3) and (4):

(1) This player made 40% of his shots last season.

(3) This player made 90% of his shots last season.

¹⁶⁹ In the next chapter, I will appeal to the same idea in discussing how an attribute frame might evoke its *counterpart* as an alternative. For now, though, I will restrict my focus to scalar alternatives.

(4) This player made 15% of his shots last season.

Meanwhile, the set of alternatives to (2) would include (5) and (6).

(2) This player missed 60% of his shots last season.

(5) This player missed 75% of his shots last season.

(6) This player missed 20% of his shots last season.

Thus far, Geurts's analysis says nothing about the significance of the scale reversal in accounting for *framing effects*; it simply identifies sets of alternatives that arise in virtue of a frame's linguistic structure. However, Geurts goes on to argue that when recipients of frames provide evaluative responses, their utterances convey frame-dependent information about items in the set of alternatives. Specifically, he suggests that when hearers evaluate a target entity *favourably*, they convey that alternatives higher up the scale would have been even better, while alternatives lower down would have been worse. To illustrate how this works, the central scenario Geurts considers concerns the following two ways of describing a plane crash, affecting 600 passengers:

(7) 200 people survived.

(8) 400 people died.

He observes that, whereas it would be felicitous to say (9) it would be infelicitous to say (10):

(9) It's good that 200 people survived.

(10) It's good that 400 people died.

His explanation is that (9) conveys that it would have been *better* if more than 200 people survived. This makes (9) felicitous, since that *would* have been better. In contrast, (10) infelicitously conveys that it would have been better had more than 400 people *died*. Geurts writes:

[E]ven if [a speaker's] descriptions of the *actual* facts are equivalent ('200 people survived/ 400 people died, and that's good'), [9] and [10] also license inferences about *counterfactual* states of affairs, i.e. about what might have been the case, and these turn out to be inconsistent (Geurts, 2013, p. 12, emphasis in original).

Geurts' reasoning reads across nicely to the example of the basketball player: giving a *favourable* evaluation of the player in response to (1) could be understood as conveying that the player would have been better had he made more shots (as per alternative (3) above, for example), and would have been worse had he made fewer shots (as per alternative (4)). In

contrast, giving an *unfavourable* evaluation in response to (1) could be understood as conveying that the player would have been worse had he made more shots, and better had he made fewer. Since that would clearly be odd, it would be an infelicitous response.

On the flipside, giving a *favourable* evaluation of the player in response to (2) could be understood as conveying that the player would have been better had he missed more shots (as per alternative (5), for example), and worse had he missed fewer (as per alternative (6)). Again, this would be odd. Evaluating the player *unfavourably*, on the other hand, would be perfectly felicitous, since it conveys that the player would have been worse had he missed more shots, and better had he missed fewer. In general, then, a favourable evaluation will be felicitous when the scale is ordered by a positively-valenced expression but not when it is ordered by a negatively-valenced one (which instead calls for an *unfavourable* evaluation).

In this way, Geurts' hearer-focused proposal could explain the framing effect in a case like that of the basketball player: recipients of the positively-worded 'made' frame will tend to evaluate the player relatively favourably, while recipients of the negatively-worded 'missed' frame will tend to evaluate him relatively *unfavourably*.

Note that Geurts takes an importantly different line from the information leakage theorists here: rather than appealing to the information typically conveyed by a speaker's frame, he focuses instead on the constraints which that frame imposes on hearers; specifically, on what they can convey with their evaluative judgements. This is an extremely interesting proposal. Taken on its own, though, I do not believe it can explain why *speakers'* framing choices are sensitive to contextual information concerning reference points, as well as *hearers'* elicited judgements. Moreover, it seems not to extend to evaluatively neutral cases of framing effects, since it depends on a contrast between positively- and negatively-valenced wording. In that respect, it is in the same boat as the associationist and implicit recommendation hypotheses.

At best, then, Geurts' proposal gives us a *partial* explanation of the framing data. To provide a full explanation, the proposal would need to be supplemented by something like the reference point hypothesis. On the one hand, that seems unproblematic, since the two accounts are *compatible*: it could be that *both* speakers' frames *and* hearers' responses convey information beyond the explicit semantic meanings of their words. Indeed, Geurts himself may be envisaging such a combined approach: he concludes his 2013 article by briefly discussing some of the other framing researchers who acknowledge that "the information conveyed by a sentence goes beyond its descriptive content" (Geurts, 2013, p. 16). Having explicitly cited the research by McKenzie and Nelson (2003), Geurts continues:

All these scholars agree that the information carried by a sentence is not exhausted by its descriptive content. Hence, there seems to be a consensus that we need a notion of information that is essentially richer than the classical one. Indeed, I believe that the consensus runs deeper, that despite obvious differences in terminology and emphasis, the ideas underlying these various approaches are closely related, and that the connections can be brought out within the general framework outlined in the foregoing. (Geurts, 2013, pp. 16-17)

On the other hand, though, the reference point hypothesis already holds out the promise of explaining the data (pertaining to both speakers' and hearers' behaviour, and both valenced and neutral frames). For this reason, in the remaining chapters of the thesis I choose to focus solely on that hypothesis.

10. Conclusion

I have argued that alternative frames plausibly convey distinct information beyond their (logically equivalent) minimal semantic meanings. According to the most theoretically powerful and empirically well-supported hypothesis on the market, alternative frames convey distinct contextual information about *reference points*. In the next chapter, I will proceed to argue that this information should be thought of as being pragmatically 'implicated'.

CHAPTER 7

Implicating Reference Point Information

Abstract

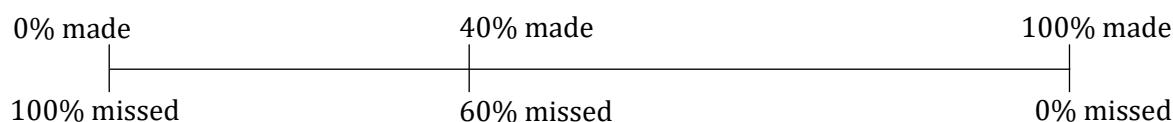
In this chapter, I explore the ‘pragmatic’ credentials of the information leakage account. I argue that frames conversationally implicate reference point information, showing that this information is not only cancellable but also calculable. I dust off an additional maxim of Manner, proposed by Grice in his ‘Presupposition and Conversational Implicature’, and explain how this licences hearers to infer reference point information from a speaker’s use of frame. I end by considering the status of reference point information as part of what speakers intend to communicate: although I acknowledge that, intuitively, this information is not always uppermost in a speaker’s mind, I argue that it still counts as something s/he implicates.

1. Introduction

In the previous chapter, I endorsed a version of the ‘information leakage’ account according to which a pair of alternative frames may be (‘explicitly’) logically equivalent to one another while ‘implicitly’ conveying distinct information – specifically, information about contextual reference points. As noted there, I believe the distinction between ‘explicit’, or ‘logical’, meaning and ‘implicit’, or ‘informational’, meaning may be helpfully cast in terms of the semantics-pragmatics divide.¹⁷⁰ First, it is somewhat misleading to contrast ‘logical’ and ‘informational’ meaning since, in whichever respect alternative frames are supposed to be logically equivalent, they will also be carrying the same information. For example, (1) and (2) carry the same information about a player, locating his performance at exactly the same point marked on the scale below:

(1) This player made 40% of his shots last season.

(2) This player missed 60% of his shots last season.



¹⁷⁰ Corner and Hahn (2010) similarly interpret the information leakage account as tracking that distinction; however, they do not engage in any detailed discussion of how the notions of semantics and pragmatics should be understood.

Indeed, Sher and McKenzie (2011) seem to recognise this point, distinguishing there between different ‘levels’ of information.

By the same token, the distinct pieces of reference point information conveyed by alternative frames are *logically inequivalent* to one another. For example, (3) and (4) clearly have different truth-conditions:

(3) This player made a relatively large proportion of his shots last season.

(4) This player missed a relatively large proportion of his shots last season.

The descriptors ‘explicit’ and ‘implicit’ are perhaps somewhat better, although these get used in different ways by different theorists.¹⁷¹ I propose, instead, to map the distinction to that of ‘semantics’ vs. ‘pragmatics’. Of course, it should be acknowledged straight off the bat that these terms also have an extremely vexed history in philosophy and linguistics. Over the course of the next two chapters, though, I will use the reference point hypothesis to argue for, and clarify, a broadly Minimalist conception of the semantics-pragmatics divide. In the current chapter I will show how the pragmatics of frames can be thought of in broadly Gricean terms. In the subsequent chapter, I will go on to argue that their semantics should be conceived of minimally.

Note that Sher and McKenzie explicitly take themselves to be putting forward a ‘pragmatic’ explanation of framing effects, understood as one which concerns “the ways in which speakers typically select utterances and convey meaning in human conversational environments” (McKenzie & Sher, 2008, p. 83). Describing their project, they write:

This chapter demonstrates how the pragmatics of natural language can explain many apparent biases of ‘framing’ in reasoning. It explains that the way in which a problem is expressed carries a great deal of information, e.g. about underlying beliefs and values – and that these influence reasoners in systematic ways, although the formulation of a problem has typically been viewed as irrelevant to its solution (McKenzie & Sher, 2008, p. 79).

Nevertheless, the authors remain deliberately neutral with respect to the particular pragmatic mechanism at work:

Note that the present analysis makes no assumptions about the existence of Gricean norms, or, more generally, about the communication of informative intent...The analysis simply points out that, when a certain kind of regularity in speaking behavior exists, a

¹⁷¹ For example, as discussed in chapter 2, Relevance Theorists associate ‘explicit’ meaning with ‘explicatures’, which include pragmatic modulation and enrichment of the conventional meanings of the words actually used. Bach (1994) rejects this approach, arguing that such effects remain *implicit*.

particular kind of inference will typically be warranted, norms and intentions aside. Whether and how listeners, in drawing such inferences, consider informative intentions or conversational norms is a question for further research to address. (Sher & McKenzie, 2006, p. 470, footnote 3)

Here I will develop their analysis by showing that the reference point information conveyed by a frame *can* be understood as a conversational implicature in Grice's sense. The structure of the chapter is as follows: in §2, I introduce the Gricean concept of conversational implicature. In §§3-5 I work through three case studies, to assess whether the reference point information inferred from frames could plausibly be a (relevance- or quantity-based) implicature. I identify three potential obstacles to such an analysis, based on reference point information being *detachable*, *unintended*, and *non-calculable*. In §6 I argue that these obstacles cannot be overcome by appeal to Grice's original set of Manner maxims, or to the related notion of 'markedness'. In §7, however, I show how an additional maxim of Manner, as proposed by Grice in 'Presupposition and Conversational Implicature', would render reference point information calculable (as well as explaining its detachability). In §8 I turn to the issue of speakers' intentions, arguing that, even if implicatures must be part of the speaker's intended meaning (which is not obvious), that would not automatically disqualify reference point information from counting as implicated.

2. Conversational Implicature

Grice famously distinguished between what is 'said' by a sentence and what it may be used to 'implicate' on a particular occasion of utterance (Grice, 1989, chapter 2).¹⁷² According to Grice, when a speaker utters a complex expression, what is 'said' will be more or less the result of composing the conventional meanings of the constituent words ('more or less' because it will also depend on the resolution of any ambiguities, and on the assignment of denotations to any context-sensitive expressions (Grice, 1989, p. 25)). Grice further required that what speakers 'say' be part of what they *mean*. His definition of speaker meaning is provided below – note that the subscript 'NN' refers to 'non-natural' meaning, as distinct from meaning that merely correlates naturally occurring events (as in smoke 'meaning' fire).

"A meant_{NN} something by x" is roughly equivalent to "A uttered x with the intention of inducing a belief by means of the recognition of this intention." (Grice, 1989, p. 219)

¹⁷² I will largely ignore *conventional* implicatures, which Grice took to depend on expressions' conventional meanings without affecting what is 'said'. Due to the cancellability of reference point information (which I will demonstrate in §3), I do not believe that information could be conventionally implicated. Throughout the chapter, then, I will use 'implicature' interchangeably with 'conversational implicature'.

Thus, information which is speaker meant (or 'm-intended') can be thought of as that which the speaker intends the hearers to recover, by means of their recognising that the speaker had that very intention.¹⁷³

In contrast with what is 'said', *implicatures* can depart quite radically from conventional meaning. This is because they depend in a substantive way on general rational principles of communication, summarised by the Cooperative Principle:

Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged. (Grice, 1989, p. 26)

Beneath this overarching principle, Grice (1989, pp. 26-27) proposed a series of conversational maxims, falling within the categories of Quantity, Quality, Relation, and Manner:

Quantity

1. Make your contribution as informative as is required (for the current purposes of the exchange).
2. Do not make your contribution more informative than is required.

Quality: Try to make your contribution one that is true.

1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

Relation: Be relevant.

Manner: Be perspicuous.

1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief (avoid unnecessary prolixity).
4. Be orderly.

These are intended to be highly intuitive principles, which are characteristic of rational behaviour in general. As Grice writes:

¹⁷³ The requirement for what is 'said' to answer to both conventional and communicational criteria generates a tension in Grice's account. For further discussion, see (Borg & Fisher, forthcoming). For the purposes of the current chapter, I set this issue aside (although it will feature again in the next chapter).

As one of my avowed aims is to see talking as a special case or variety of purposive, indeed rational, behavior, it may be worth noting that the specific expectations or presumptions connected with at least some of the foregoing maxims have their analogues in the sphere of transactions that are not talk exchanges. (Grice, 1989, p. 28)

He observes that similar principles apply to non-linguistic activities like mending a car or baking a cake. The common sense nature of the conversational maxims forms part of their intuitive appeal: as Geurts (2010, p. 11) puts it, “the banality of the maxims strongly speaks in their favour”.

2.1. Calculability

Grice argued that, on the assumption that speakers are complying with the maxims (or, at least, the Cooperative Principle), it is possible to ‘calculate’ that a speaker is implicating something beyond what they merely ‘say’. To illustrate the calculability of implicatures, Grice invites us to consider the following exchange:

Suppose that A and B are talking about a mutual friend, C, who is now working in a bank. A asks B how C is getting on in his job, and B replies, *Oh quite well, I think; he likes his colleagues, and he hasn't been to prison yet.* (Grice, 1989, p. 24)

He provides the following gloss to show how B implicates that C is potentially dishonest:

In a suitable setting A might reason as follows: “(1) B has apparently violated the maxim ‘Be relevant’ and so may be regarded as having flouted one of the maxims conjoining perspicuity, yet I have no reason to suppose that he is opting out from the operation of the Cooperative Principle; (2) given the circumstances, I can regard his irrelevance as only apparent if, and only if, I suppose him to think that C is potentially dishonest; (3) B knows that I am capable of working out step (2). So B implicates that C is potentially dishonest.” (Grice, 1989, p. 31)

In this example, the maxim of Relation can be used to work out that A believes C to be potentially dishonest. That information clearly goes beyond what A ‘says’, which is merely that C *hasn't been to prison yet.*

Grice takes it to be a *necessary* condition of some content’s being implicated that it be calculable from the conversational principles. He writes:

A man who, by (in, when) saying (or making as if to say) that *p* has implicated that *q*, may be said to have conversationally implicated that *q*, provided that (1) he is to be presumed to be observing the conversational maxims, or at least the Cooperative

Principle; (2) the supposition that he is aware that, or thinks that, q is required in order to make his saying or making as if to say p (or doing so in *those* terms) consistent with this presumption; and (3) the speaker thinks (and would expect the hearer to think that the speaker thinks) that it is within the competence of the hearer to work out, or grasp intuitively, that the supposition mentioned in (2) is required. (Grice, 1989, pp. 30-31, emphasis in original)

Unpacking this rather complex passage, the first clause specifies a condition known as the ‘cooperative presumption’: it must be assumed that the speaker is adhering to the conversational maxims (or, at least, to the Cooperative Principle) in order to work out the implicature. The second clause specifies a condition I will call the ‘calculability criterion’. This states that the cooperative presumption must require a belief to be attributed to the speaker, the content of which goes beyond what was ‘said’. In this sense, the implicature is needed to bridge the gap between what a speaker ‘said’ and the assumption of the speaker’s cooperativity. Finally, the third clause requires that the speaker believes that hearers can, in fact, infer that the implicature is required.

I want to highlight four aspects of calculability that will be important in the subsequent discussion. Two can be dealt with fairly quickly here; the other two will be addressed in §§2.2-2.3. The first point to note is that Grice required implicatures to be calculable whether they are ‘particularised’ (only being conveyed in virtue of special features of the context) or ‘generalised’ (normally being conveyed, in the absence of special contextual features) (Grice, 1989, pp. 37-38). He writes:

The presence of a conversational implicature must be capable of being worked out; for even if it can in fact be intuitively grasped, unless the intuition is replaceable by an argument, the implicature (if present at all) will not count as a conversational implicature; it will be a conventional implicature. (Grice, 1989, p. 31)

This point will be relevant in §7, where I discuss the relatively generalised way in which frames convey reference point information.

Second, it is worth emphasising that the calculability criterion makes no claims about how implicatures are actually processed in the minds of interlocutors (this point is emphasised, for example, by Bach (2006b)). There is ongoing debate in the philosophical literature about whether a Gricean calculation process is psychologically realistic – see, for example (Geurts & Rubio-Fernández, 2015; Saul, 2002b). However, this is a point on which I intend to remain neutral throughout. Two further issues require a little more discussion.

2.2. Indeterminacy

One point concerns the notion of *requirement* in the above-quoted passage from Grice. The claim that it should be possible to calculate some particular content ('q') seems too strong, even by Grice's own lights. A few pages later, he notes that the contents of implicatures are often indeterminate:

Since, to calculate a conversational implicature is to calculate what has to be supposed in order to preserve the supposition that the Cooperative Principle is being observed, and since there may be various possible specific explanations, a list of which may be open, the conversational implicatum in such cases will be [a] disjunction of such specific explanations; and if the list of these is open, the implicatum will have just the kind of indeterminacy that many actual implicata do in fact seem to possess. (Grice, 1989, pp. 39-40)

I suggest, then, that we should not think of hearers as being forced to attribute a specific belief to the speaker. However, hearers are still required to suppose that the speaker has *some* extra belief. In other words, the cooperative presumption determines that an implicature is *present*, even if does not determine what it *is*.

The same point emerges from a dispute between Davis (1998) and Saul (2001). Davis argues that many apparent implicatures are not calculable, but depend instead on linguistic conventions.¹⁷⁴ He re-examines standard Gricean analyses of tautologies like 'war is war' (Davis, 1998, p. 41). A speaker of this sentence would appear to violate the first maxim of Quantity, since what is 'said' is totally uninformative. Therefore, if the speaker is to be understood as cooperative, s/he must be presumed to have something else in mind. However, Davis points out that the conversational principles give us no way to arrive at what that something else might be. Instead, it is only due to particular linguistic conventions of English that speakers may use the expression 'war is war' to communicate something like 'It is the nature of war that terrible things happen' (Davis, 1998, pp. 43-46/ 87-89).¹⁷⁵

According to Davis, tautology implicatures fail to meet Grice's calculability criterion. Against this, Saul (2001) argues that Grice's framework can accommodate such implicatures. After all, Grice himself was clear that the cooperative presumption is only one of the factors on which implicatures depend:

¹⁷⁴ As discussed in chapter 2, a similar idea has been defended more recently by Lepore and Stone (2015).

¹⁷⁵ Davis draws on the cross-linguistic analysis of Wierzbicka (1985), which shows that the contents of tautology implicatures are contingent on the conventions of particular languages.

To work out that a particular conversational implicature is present, the hearer will rely on the following data: (1) the conventional meaning of the words used, together with the identity of any references that may be involved; (2) the Cooperative Principle and its maxims; (3) the context, linguistic or otherwise, of the utterance; (4) other items of background knowledge; and (5) the fact (or supposed fact) that all relevant items falling under the previous headings are available to both participants and both participants know or assume this to be the case. (Grice, 1989, p. 31)

Building on this, Saul sees no reason why the 'background information' Grice refers to shouldn't include knowledge of community-wide conventions. She writes:

Once this information is allowed to enter into the audience's calculations, tautology implicatures become perfectly calculable. With this information a part of the calculation, audiences also become much more restricted as to what hypotheses are reasonable to consider in trying to make sense of the speaker as cooperative. Given the background information about historical conventions, it seems far more plausible to suppose that the audience might be required to assume the conventional tautology implicatures. (Saul, 2001, p. 638)

Saul's interpretation of Grice here supports my claim that the conversational principles might determine an implicature's *presence* without necessarily determining its *content*. This point will come up again in §4.

2.3. Speaker Meaning

It is generally assumed that implicatures must be part of speaker meaning, in Grice's technical sense of m-intending. However, this is not inevitable. Certainly, no requirement for implicatures to be speaker-meant falls out straightforwardly from Grice's three necessary conditions, quoted earlier. In particular, just because the speaker believes that the hearer *can* infer the presence of an implicature does not guarantee that the speaker *intends* the hearer to do so. Demonstrating this, Saul (2002a, pp. 237-238) considers a case in which a professor is providing a reference for a candidate whom she believes to be a terrible philosopher but likes on a personal level. The professor writes a long and irrelevant (yet factually accurate) reference in the hope that the search committee, despite being *capable* of working out that the candidate is a terrible philosopher, will in fact fail to do so.

Of course, though, Grice only provides us with *necessary* conditions for implicating. These need not be jointly *sufficient*. Grice could therefore be assuming that there is a further requirement for implicatures to be speaker-meant. This is how he is interpreted, for example, by Davis

(1998) and Neale (1992). Again, though, Saul (2001, 2002a) provides a more nuanced interpretation, observing:

[I]t is genuinely unclear whether Grice took being meant by the speaker to be a necessary condition for being a conversational implicature. He may well have done so, but the only place in which he states this explicitly is a passage in “The Causal Theory of Perception” which he chose to omit when he collected his papers for *Studies in the Way of Words*. (Saul, 2001, p. 635)

Although I cannot settle the exegetical debate here, I return to this point in §8, in discussing the worry that reference point information is not speaker-meant.

2.4. Cancellability and Nondetachability

I conclude the section on Gricean implicature by describing two other features of implicatures, in addition to their calculability. First, conversational implicatures are characteristically *cancellable*, whereas what is ‘said’ is not (Grice, 1989, p. 39). For example, imagine a modification of the earlier exchange, in which B responds as follows to A’s question about how C is getting on in his job:

(5) Oh quite well, I think; he likes his colleagues, and he hasn’t been to prison yet. I was quite worried about that since the last three jobholders were framed for fraud.

B’s continuation here explicitly pre-empts the implicature that C is dishonest. Implicatures can also be defeated implicitly, by wider contextual factors. For example, if both A and B knew (and knew that the other knew) that the last three jobholders were framed for fraud, then the implicature that C is dishonest is unlikely to arise.

Related to their cancellability, Sadock (1978) observes that conversational implicatures can be reinforced without redundancy. Thus, for example, B might have felicitously uttered (6):

(6) Oh quite well, I think; he likes his colleagues, and he hasn’t been to prison yet, despite his being dishonest.

Here, explicitly appending the information that C is dishonest to the original utterance seems perfectly acceptable.

Grice also observed that implicatures are typically ‘nondetachable’, meaning that they survive logically equivalent reformulations of what is ‘said’ (Grice, 1989, p. 39). Thus, for example, had B said something like the following, he would still have implicated that C is potentially dishonest.

(7) Oh quite well, I think; he likes his colleagues, and he hasn’t ended up in the nick yet.

An important qualification here is that Manner-based implicatures will be ‘detachable’, since they are linked to the way in which something is said, not just *what* is said. This will be important in §7, where I propose that reference point information is a Manner implicature.

3. Reference Point Information and Relevance

I turn now to consider whether the reference point information conveyed by a frame is analysable as a conversational implicature. In a relatively straightforward type of case, the information may be thought to arise from considerations of *relevance*. Consider, for example, the following hypothetical exchange concerning a basketball player:

A: Why wasn’t Federico recruited to the team?

B: He missed 60% of his shots last season.

Assuming that B is cooperative, Federico must have missed a *relatively large* proportion of shots, otherwise B’s utterance would fail to provide a relevant answer to A’s question.

In addition to its being calculable in this way, the reference point information conveyed by B’s utterance is also cancellable. For example, B might have continued as follows:

(8) He missed 60% of his shots last season. That’s a relatively small proportion of misses but, unfortunately, the coach witnessed every single one.

In (8) B explicitly blocks the reference point information, defeating the implicature that might otherwise be present.

Relatedly, no redundancy would result from making the reference point information explicit, as in (9):

(9) He missed 60% of his shots last season. That’s a lot!

Note that the fact that reference point information is readily defeated or reinforced indicates that it does not arise purely as a result of composing the conventional meanings of the words used. Instead, it requires modulation or enrichment of some sort.¹⁷⁶

Even in this relatively straightforward case, however, we encounter some problems with the nondetachability criterion. On one hand, the reference point information survives certain reformulations. For example, had B uttered (10) she would still have implicated that Federico missed a relatively large proportion of his shots:

¹⁷⁶ In the next chapter, I will consider whether it could nevertheless intrude on what is ‘said’ in some sense.

(10) 60% of his shots failed to go in last season.

On the other hand, as predicted by the reference point hypothesis, it may not be possible to convey the reference point information when the alternative frame is used. Consider the following version of the exchange:

A: Why wasn't Federico recruited to the team?

B: He made 40% of his shots last season.

B's utterance sounds odd here and is difficult to square with the assumption that she is providing a relevant answer to A's question.¹⁷⁷ I will end up diagnosing this as a conflict between the maxims of Relation and Manner (see §7). For now, I want to introduce a trickier case.

4. A Calculability Gap?

Recall that, in the experimental scenario used by McKenzie and Nelson (2003), which was discussed in the previous chapter, the protagonist, Mary, describes a glass by using a 'full' or 'empty' frame. The example vignette is reproduced below:

Imagine that Mary was sitting at her kitchen table with a glass in front of her. She left the room briefly and came back to find that the contents of the glass had changed. When asked to describe the glass now, Mary said, "The glass is half full." Given how Mary chose to describe the glass after its contents had changed, please choose the statement below in terms of what you think was most likely true about the glass before its contents changed.

The glass was full before its contents changed.

The glass was empty before its contents changed.

As discussed, McKenzie and Nelson find that when Mary uses 'full' frames, participants are more likely to judge that the glass was previously *empty* than when she uses 'empty' frames (and, of course, vice-versa).

I want to allow that this result could potentially be given a relevance-based analysis. Note first that the experimenters ask participants to judge the prior state of the glass without giving them

¹⁷⁷ Notice that the inclusion of 'only' – i.e. 'He *only* made 40% of his shots last season' – would restore the original implicature. Thus, 'only' switches the reference point from making fewer shots to making more. It may be possible to achieve a similar effect with certain patterns of intonation. I take this as further evidence that frame-dependent reference point information is usually present in the absence of such features, and that it can be difficult to dispel, even in light of considerations of relevance.

the requisite information to *deduce* an answer. Merely composing the conventional meanings of Mary's words will not suffice, since the result is strictly compatible with the glass having been full or empty. Nevertheless, participants are instructed to consider how Mary chose to describe the glass. Perhaps, then, the utterance the experimenters put in Mary's mouth should be thought of as implicating *something* about the prior state of the glass (even if it is unclear *what*). After all, the experimental stimuli would otherwise be irrelevant to the task, violating the maxim of Relation (or, alternatively, the stimuli might be thought of as insufficiently informative, violating the first Quantity maxim).¹⁷⁸

Note that, unlike in the exchange about Federico, appealing to Relation (or Quantity) could at most explain the *presence* of an implicature. The *content* of the reference point information cannot be derived because those maxims tell us nothing about why hearers tend to associate Mary's 'empty' frames with the glass being previously full and her 'full' frames with its being previously empty. In principle, the opposite associations would be equally relevant and informative. As discussed in §2.2, though, this may not be a problem so long as the conversational principles need only require *something* to be implicated, without their necessarily determining *what*.

A separate worry that emerges with this scenario is that it seems less obvious that Mary intends to communicate the reference point information (or, perhaps better, that the experimenters intend her to have such an intention). I will return to this issue in §8. For now, I want to present a final case, in which reference point information might be thought to be entirely non-calculable.

5. A Calculability Failure?

I will focus on the series of 'action-oriented' experiments run by Sher and McKenzie (2006), to corroborate and extend the findings of McKenzie and Nelson (2003). Below is a description of their first experiment:

Subjects, tested individually, were seated at a desk in one of two small rooms. At the left edge of the desk, two apparently identical transparent plastic cups stood side-by-side. One was full of water, the other empty. (Order of the cups—whether the full or empty cup was closest to the subject—was varied between the rooms.) Indicating the right edge of the desk, the experimenter said to about half of the subjects in each room:

¹⁷⁸ This case is complicated by the fact that the experimenters deliver information to participants via a fictional protagonist. Ultimately, a Relation-based analysis may not be workable here but that would merely expedite my later argument for a Manner-based approach.

“Just to get things started, could you pour water from one cup to the other and set a half-full cup at the edge of the desk”.

The remaining subjects were asked for “a half-empty cup”.

(Sher & McKenzie, 2006, p. 473)

Two further experiments used the following fractions in the instruction: ‘one-quarter full’ or ‘three-quarters empty’ (in the second experiment); and ‘three-quarters full’ or ‘one quarter empty’ (in the third experiment).¹⁷⁹

The experimenters found that participants’ choice of which cup to set at the edge of the desk varied systematically across conditions, depending on whether the instructor used a ‘full’ or ‘empty’ frame. Specifically, participants who received ‘full’ frames were more likely to choose the initially *empty* cup than those who received ‘empty’ frames (and vice-versa). The full results are shown in the table below. As in McKenzie and Nelson’s study, there was no significant difference between the ‘one-quarter empty’ and ‘three-quarters full’ conditions, although the trend was in the same direction.

| Frame | % participants choosing the glass that was <i>full</i> before | % participants choosing the glass that was <i>empty</i> before |
|---------|---|--|
| ¼ full | 7% | 93% |
| ¾ empty | 29% | 71% |
| ½ full | 46% | 54% |
| ½ empty | 69% | 31% |
| ¾ full | 60% | 40% |
| ¼ empty | 70% | 30% |

¹⁷⁹ In these experiments, the set up was slightly different:

As in Experiment 1, subjects were seated at a desk in one of two small rooms. Transparent cups, one full and one empty, stood side-by-side at the left edge of the desk. (Again, order of the cups was varied between rooms.)

A square, slightly larger than the base of a single cup, was marked on the desk. Subjects were given a one-page instruction sheet, which, for half of the Experiment 2 subjects, read:

In front of you on the table you’re sitting at, there should be two cups and a square. To get things started, please pour water from one cup to the other and set a 3/4-empty cup in the square. Please go tell the experimenter when you’ve finished, and he will tell you what to do next.

The remaining subjects in Experiment 2 were asked for “a 1/4-full cup” instead of “a 3/4-empty cup”.

In Experiment 3, the instruction sheet, otherwise identical to that used in Experiment 2, requested either “a 1/4-empty cup” or “a 3/4-full cup”.

(Sher & McKenzie, 2006, p. 475)

The experimenters note that the use of written rather than spoken instructions eliminates certain possible confounds, such as prosodic effects. It seems unlikely that the other differences (such as asking the participant to place the cup in the square, rather than at the edge of the desk) affect comparability.

The results of these experiments indicate that participants remained sensitive to reference point information concerning the prior contents of the cup. Crucially, though, in this experiment, the prior state of the cup is strictly *irrelevant* to the completion of the task. What the experimenters ‘say’ is already entirely consistent with their being cooperative speakers. They are not obliged to care *which* of the two cups is placed at the edge of the desk, just that one of them is. Thus, participants are given the freedom to choose how they pour the water between the cups.¹⁸⁰ After all, the experimenters might have been interested, say, in which hand the participant would use, or how close to the edge of the desk the cup would be placed. Indeed, in the second and third experiments, participants were asked to judge the probable purpose of this part of the experiment and it appeared to have remained opaque to them. Here, then, it is implausible that there could be any relevance- or quantity-based implicature.

Sher and McKenzie’s results suggest that reference point information can be inferred without its being implicated via the maxims of Relation or Quantity. Perhaps, at this point, we might try appealing to Manner – after all, what is important is *how* the speakers framed their utterances. I think it would be difficult to directly apply Grice’s original set of Manner maxims; these are designed to deal with linguistic formulations that are especially obscure, ambiguous, prolix, or disordered – or, as Lepore and Stone (2015, p. 104) put it, “so cumbersome or awkward that no speaker could sensibly use it without some ulterior motive”. In contrast, the attribute frames considered here, using ‘full’ or ‘empty’, ‘made’ or ‘missed’, seem clear and straightforward. That said, there might be scope to argue that one member of the pair is more ‘marked’ than its counterpart. In the next section I explore how this proposal could work.

6. Markedness

The notion of ‘markedness’ concerns the morphological complexity or unusualness of an expression. As Levinson writes:

On the formal side, marked forms, in comparison to corresponding unmarked forms, are more morphologically complex and less lexicalized, more prolix or periphrastic, less frequent or usual, and less neutral in register. On the meaning side, such forms suggest some additional meaning or connotation absent from the corresponding unmarked forms. (Levinson, 2000, p. 137)

¹⁸⁰ Note too the use of indefinite descriptions, as in ‘a half-full cup’, which avoids any supposition that the speaker has a particular cup in mind.

It should be noted that the phenomenon of markedness has been developed in various different ways. Lehrer (1985) surveys a number of tests that have been proposed for assessing which member of a pair of antonyms is ‘marked’ and which is ‘unmarked’. For example, unmarked expressions typically sound more natural and neutral than their marked counterparts in certain question forms. Thus, the relative naturalness of ‘How tall is Federico?’, as opposed to ‘How short is Federico?’ (which carries the additional suggestion that Federico *is* relatively short), demonstrates that ‘short’ is marked relative to ‘tall’.

I will not argue here for any particular set of criteria for identifying (un)marked expressions. Instead, I want to consider whether an asymmetry in the markedness of alternative frames could, in principle, explain how reference point information is being implicated. Focusing on the ‘full’ and ‘empty’ frames used in the experiments discussed above, I will simply follow others in the literature in assuming that ‘full’ is unmarked, whereas ‘empty’ is relatively marked (Holleman & Pander Maat, 2009; Ingram et al., 2014; Moxey, 2011).¹⁸¹ Justifying their classification, Holleman and Pander Maat (2009, p. 2210) note that “[t]his seems to be supported by the fact that *full* tends to precede *empty* when the two terms are listed, and by the use of *full* in neutralized contexts such as *how X is it?*”.

Assuming that ‘empty’ frames are marked, then, we might hypothesise that *only these frames* carry reference point information, while ‘full’ frames carry no reference point information at all beyond their conventional meanings. On this hypothesis, the framing effect is driven by the fact that ‘empty’ licenses the inference to relative emptiness, just as ‘How short is Federico?’ licenses the inference to his relative shortness. Although this is an interesting proposal, there is some empirical evidence to suggest it is unlikely to be correct. Below, I briefly summarise a study by Ingram et al. (2014).

Ingram et al. used an eye-tracking paradigm to measure the ease with which reference point information was processed under alternative frames.¹⁸² Using a within-subjects design, Ingram et al. presented participants with 36 scenarios (interspersed with 36 filler items), an example of which is provided below:

¹⁸¹ If, instead, ‘full’ and ‘empty’ are considered to be equally unmarked, then it is already clear that the proposal discussed in this section cannot explain how reference point information is implicated. I will take a few more steps to reach that same conclusion below.

¹⁸² Roughly, it was assumed that participants had more difficulty processing a region of text if they spent longer fixating on that region (allowing for differences in word length between different conditions), or if they looked back over earlier regions.

Summer was here, and Henry was eagerly preparing to get into his paddling pool. He ran inside to ask his Mum to apply sunscreen then hurried back to the garden. The pool was now *half full* but when he had left it was *full* and the sun beat down.¹⁸³

The text shown in italics was varied to generate the four conditions in the table below.

| Frame | Prior state |
|--------------|-------------|
| 'Half full' | Full |
| 'Half full' | Empty |
| 'Half empty' | Full |
| 'Half empty' | Empty |

In two further control conditions, the critical expressions in the frames were replaced by expressions of similar length that were unrelated to volume, such as 'quite cold'. The prior state was still described as 'full' or 'empty' in these conditions. The purpose of the controls was to see how easily the prior state would be processed when no reference point information could have been conveyed.

The experimenters found framing effects on processing: when a container's current state was described using a 'full' frame, participants were quicker to process information about its prior state if it had previously been empty rather than full; and, conversely, when the container's current state was described using an 'empty' frame, participants were quicker to process information about its prior state if it was previously full rather than empty. In other words, readers seemed to be expecting a prior state of 'full' after the use of an 'empty' frame and – crucially – a prior state of 'empty' after the use of a full frame. The results suggest that the framing effect is not being driven solely by inferences drawn from 'empty' frames but also by those drawn from 'full' frames. Ingram et al. conclude:

In this experiment, participants' patterns of reading strongly reflected the predictions of the information leakage framework. A consistent interaction within the critical region showed that where a vessel was initially described as being *half full*...there was a clear processing advantage for a previous volume of *empty* (Ingram et al., 2014, p. 814).¹⁸⁴

¹⁸³ The purpose of the final clause is to allow gaze to be tracked after the critical region.

¹⁸⁴ One might worry that the corresponding processing delay for a previous volume of *full* in this condition was simply due to the oddness of repeating the expression 'full'. This is not addressed by Ingram et al. and would be a useful question for further empirical research to pursue.

While it would certainly be valuable to confirm this result using other pairs of frames and different experimental paradigms, I will assume for now that framing effects are not driven *purely* by marked forms.¹⁸⁵

Before moving on, I will briefly address – and reject – another possibility. One thought at this point might be that *both* frames convey reference point information but that the unmarked ‘full’ frame conveys how glasses *typically* relate to reference points, whereas the marked ‘empty’ frame conveys some unusual relation between the glass and a reference point. However, reflection quickly shows that this proposal is untenable. It would require that glasses are *typically* relatively full (that being the reference point information associated with the unmarked ‘full’ frame). Even if that supposition were credible, the logic certainly couldn’t carry across to other cases. For example, returning to the study by Leong et al. (2017), it would be incoherent to suppose that basketball players *typically* make (or miss) a relatively large proportion shots – by definition, they typically make (and miss) the typical proportion!

In sum, it is doubtful that reference point information is calculable purely on the basis of markedness.¹⁸⁶ That said, I do think marked forms could convey reference point information with greater *strength* than unmarked forms. I return to this idea in §7.2 after first setting out the core of my positive proposal.

7. A New Manner Maxim

I will argue that reference point information can be analysed as an implicature once we put in place an additional submaxim of Manner. Initially that might seem like an ad hoc move. First, though, it should be noted that directly after introducing the four submaxims of Manner, Grice explicitly states that “one might need others” (Grice, 1989, p. 27). Second, in later work, Grice proposed an additional manner maxim himself, the wording of which is highly suggestive for current purposes:

I would be inclined to suggest that we add to the maxims of Manner which I originally propounded some maxim which would be, as it should be, vague: “Frame whatever you say in the form most suitable for any reply that would be regarded as appropriate”; or, “Facilitate in your form of expression the appropriate reply.” (Grice, 1989, p. 273)

¹⁸⁵ I note also that such a hypothesis would be unable to explain framing effects that arise where neither form is marked. For example, a series of experiments reported by Teigen and Karevold (2005) contrast time ‘spent’ or ‘left’, or work ‘done’ or ‘left’. There does not seem to be any obvious difference in the markedness of these (at least not in the English translation of the original Norwegian) yet framing effects are still observed (with hearers making systematically different inferences about progress relative to the original schedule).

¹⁸⁶ Although it would also be worth making sure that framing effects are not exhausted by the combined effects of markedness and lower-bounded representations of quantities, discussed in chapter 3.

Of course, it must be acknowledged that Grice introduces this ‘framing maxim’ for the entirely different purpose of defending a Russellian semantic analysis of statements like (11):

(11) The king of France is bald.

Grice takes (11) to entail (not presuppose) that there is a king of France. However, he argues that it carries an implicature, which is calculable on the basis of the framing maxim. Specifically, Grice suggests that it is only felicitous to assert (11) if an appropriate response would be to challenge the king’s baldness but *not* his mere existence. Grice’s rationale here is that a cooperative speaker will only convey one challengeable claim at a time and, since the king’s baldness depends on his existence whereas his existence does not depend on his baldness, it must be the baldness claim which is challengeable. If, instead, the king’s existence should also be open to challenge, then the speaker of (11) would have violated the framing maxim. S/he should instead have used a formulation like ‘There is a king of France and he is bald’.

Although I make no comment on Grice’s analysis of sentences like (11), I want to suggest that his framing maxim can account for the presence of reference point information. I propose reformulating it slightly as follows:

Framing Maxim: Facilitate in your form of expression the appropriate response.

The idea is this: if a speaker is complying with the framing maxim, s/he will choose the frame that facilitates the response of inferring accurate reference point information. In the next subsection, I will say something about what makes the inference of reference point information an ‘appropriate’ response to an attribute frame. In particular, I will argue that interlocutors have tacit knowledge of an association between frames in use and contextual reference points.

7.1. Tacitly Known Use Properties

The psychological data clearly demonstrate that attribute frames do facilitate hearers’ inferences of reference point information. Thus, when a glass is described using a ‘full’ frame, hearers have a tendency to infer that the glass is relatively full; and when it is described using an ‘empty’ frame, they have a tendency to infer that it is relatively empty. As the information leakage theorists propose, this seems to be due to hearers (tacitly) knowing that the use of a frame is sensitive to facts about reference points: when an entity instantiates some property (like fullness or emptiness) to a relatively high degree compared to the contextually salient reference point, speakers tend to use the corresponding (‘full’ or ‘empty’) frame.

As mentioned in the previous chapter, the claim that speakers’ framing choices are sensitive to contextual reference points is empirically well-supported. To give a flavour of the findings, I will now summarise one of the other experiments reported by McKenzie and Nelson (2003). Below

is an example vignette, which was presented to participants in one of six experimental conditions:

Imagine a 4-ounce measuring cup in front of you that is completely filled with water up to the 4-ounce line. You then leave the room briefly and come back to find that the water is now at the 2-ounce line. What is the most natural way to describe the cup now?

Participants were then then given the following options:

The cup is $\frac{1}{2}$ full.

The cup is $\frac{1}{2}$ empty.

In two other conditions, the water now in the cup was described as being at the 1-ounce line, or at the 3-ounce line. These three conditions, in which the measuring cup started out completely *full*, were mirrored by three conditions in which it started out completely *empty*.

The experimenters found that, where the measuring cup had started out *empty*, participants were more likely to choose ‘full’ frames to describe the cup than where it had started out *full* (and, of course, vice-versa). The difference was not significant where the water was now at the 3-ounce line, although the trend was in the same direction.¹⁸⁷ The results are shown in the table below:

| Current state | Prior state | % participants choosing ‘full’ frame | % participants choosing ‘empty’ frame |
|----------------------|--------------------|---|--|
| 1-oz line | Full | 56% | 44% |
| 1-oz line | Empty | 97% | 3% |
| 2-oz line | Full | 31% | 69% |
| 2-oz line | Empty | 88% | 12% |
| 3-oz line | Full | 75% | 25% |
| 3-oz line | Empty | 88% | 12% |

The data suggest a tendency for participants to choose frames according to their beliefs about the contextually salient reference point. As mentioned previously, this finding has been confirmed in several other speaker-focused experiments, using various pairs of frames (Honda

¹⁸⁷ It is possible that, as the glass approaches fullness, the unmarked ‘full’ frame becomes the obvious choice for speakers, regardless of prior state; in the next subsection, I discuss the interaction between reference points and markedness a little further.

& Yamagishi, 2017; McKenzie & Nelson, 2003; Sher & McKenzie, 2006; Teigen & Karevold, 2005).

In summary, then, when hearers infer reference point information, they are plausibly deploying knowledge of how speakers use frames in different contexts. In other words, frames convey reference point information in virtue of facts about how speakers tend to use frames. I want to remain neutral about the status of the association between frames and contextual reference points: perhaps it depends on an arbitrary linguistic convention (of English and, presumably, several other languages, given that studies have been conducted in Dutch, Norwegian, and Japanese, at least). Alternatively, perhaps it could be a universal fact about human cognition that we tend to highlight what is relatively abundant.¹⁸⁸ While this would be an interesting issue for future (cross-linguistic and psychological) research, I will not attempt to pursue it here. Instead, I will simply assume the existence of some kind of regularity in use, which connects alternative frames with distinct reference point information.

Given such a regularity, the appropriate response to a speaker's frame will typically be to infer the associated reference point information. And cooperative speakers can be expected to use frames accordingly. Therefore, if a cooperative speaker describes a glass using a 'full' frame, hearers may infer, on the basis of the framing maxim, that it is relatively full; and, if the glass is described with an 'empty' frame, hearers may infer that it is relatively empty. If it subsequently turned out that the reference point information was *false*, that would constitute a violation of the framing maxim, other things being equal: the frame would have facilitated an *inappropriate* response. According to the proposed analysis, then, the cooperative presumption makes reference point information calculable as a Manner-based implicature. In particular, the implicature depends on the framing maxim, taken together with a tacitly known association between frames and reference points.

I want to make a series of observations about the proposed account. First, since reference point information is taken to be a Manner-based implicature, this immediately explains why it is *detachable* (as discussed in §3). Second, because the implicature is driven by the association between frames and contextual reference points, it is likely to be fairly *generalised*. This would explain why framing effects can be felt, intuitively, even when sentences are presented out of context. Nevertheless, since the framing maxim will need to be weighed up with all kinds of other contextual factors, reference point information remains *defeasible*. For example, as

¹⁸⁸ This possibility might be a better fit with the view of implicatures Geurts puts forward when he writes: "Conversational implicatures are, first and foremost, *non-conventional*; they are not due to linguistic conventions of any kind" (Geurts, 2010, p. 9, emphasis in original) It is worth noting, though, that any convention associating frames with reference point information would have to be a convention of *use*, not a convention of *meaning*. I will argue in the next chapter for the importance of this distinction.

discussed in the previous chapter, where a hearer possesses conflicting knowledge, reference point information may not be taken up, thus mitigating any framing effect. Likewise, Levin et al. (1998, p. 164) discuss ceiling and floor effects, whereby attribute framing effects tend to be overridden by other considerations at proportions close to 0% or 100%. For example, where a basketball player has made only a negligible proportion of his shots, he might still be evaluated unfavourably, regardless of whether his performance is described in terms of the proportion ‘made’ or the proportion ‘missed’.

Third, it is interesting to observe that the association between frames and reference points does not only explain the *presence* of an implicature but also constrains its *content* (making framing cases unlike the ‘war is war’ tautology implicature considered earlier). For example, the association between the use of ‘full’ frames and relative fullness means that cooperative speakers using ‘full’ frames will not just be implicating *something* but, specifically, the information that the container is *relatively full*. In general, where the presence of an implicature depends on a regularity of use, we might expect its content to be similarly constrained. Indeed, the foregoing analysis suggests that Gricean pragmatics may be able to capture effects that some have sought to analyse in different ways. For example, Lepore and Stone argue for a non-Gricean conception of pragmatics, which they describe as follows:

It doesn’t appeal to anything like Grice’s Cooperative Principle, and typically exploits shallow cues rather than deep inferences about the speaker’s mental state. (Lepore & Stone, 2015, p. 265)

It seems possible that at least some of the ‘shallow cues’ they have in mind could be captured within a broadly Gricean framework, in a similar way as suggested here for framing cues.

Fourth, the proposed account is *metalinguistic* in the sense that the hearer must register (presumably sub-consciously) the speaker’s choice of linguistic form; this falls out directly from the framing maxim.¹⁸⁹

Finally, if, as suggested, competent interlocutors tacitly know the association between frames and reference points, then it seems reasonable to suppose that speakers (tacitly) believe that hearers can infer the reference point information. The proposal would thereby satisfy the third of Grice’s necessary conditions for conversationally implicating (as set out in §2.1). Accordingly, we shouldn’t treat reference point information as a mere downstream ‘perlocutionary’ effect of

¹⁸⁹ It is worth emphasising that the linguistic alternatives in question here are not *scalar* alternatives of the kind discussed in the previous chapter. While the latter have *asymmetrical* entailment relations, frames are *mutually entailing*, at least in terms of their minimal semantics. Nevertheless, I think it is plausible that alternative frames could interact with features like prosodic focus, in similar ways as scalar alternatives. Again, for reasons of space, this it is not something I am able to discuss further here.

the utterance but as a genuine part of meaning. By way of contrast, consider a case in which A says loudly that B will play basketball tonight. Hearers might infer that B is tall on the basis that basketball players usually are. Hearers might also infer that A has a loud speaking voice. However, there is no reason to think these inferences are tracking A's beliefs; they are simply knock-on effects of the utterance. In contrast, reference point information *does* reliably track speakers' beliefs.

7.2. Interaction with Markedness

In this section, I explain why a marked form may convey reference point information more robustly than an unmarked form, playing a larger role in producing framing effects.¹⁹⁰ To illustrate the point, I will continue to use the scenario of a glass described as part 'full' or part 'empty'.

In contexts where a glass is, in fact, relatively full, it is clear how a cooperative speaker should behave: namely, by using a 'full' frame to describe the glass. First, this frame would provide accurate information about the reference point. Second, it is also the unmarked form. Thus, both considerations mutually reinforce the choice of a 'full' frame (and make an 'empty' frame particularly inappropriate). That situation is captured in the first row of the table below.

| Context | Reference-Point Sensitive Frame | Unmarked Frame |
|---------------------------|---------------------------------|----------------|
| Glass is relatively full | 'Full' | 'Full' |
| Glass is relatively empty | 'Empty' | 'Full' |

Contexts in which a glass is relatively *empty* are more complicated, however. A cooperative speaker faces the following dilemma: on one hand, the contextual reference point calls for the use of an 'empty' frame. Nevertheless, the speaker has a conflicting reason to use the unmarked 'full' form, since it is less effortful to produce, easier to understand, and so on. In this situation, then, Manner-based considerations may pull in opposite directions, as shown in the bottom row of the table below.

An upshot is that, when a marked frame *is* used, it can be expected to convey the associated reference point information with greater strength than an unmarked frame. After all, the marked form is felicitous only in one kind of context. In the table above, the use of the 'empty'

¹⁹⁰ Holleman and Pander Maat (2009) make a similar proposal in the context of their 'argumentative orientation' account, which was discussed in the previous chapter. They claim that the effects of argumentative orientation are modulated by markedness, with marked forms strengthening the effects and unmarked forms weakening them.

frame only corresponds to contexts in which the glass is emptier than before. In contrast, the unmarked form is less informative, since it can be used relatively felicitously in both kinds of context. Thus, in the table above, the ‘full’ frame can be used in contexts in which the glass is fuller *or* emptier than before. Another way of thinking about this is that the use of the ‘empty’ frame requires more justification for the extra effort required in its processing, strongly implying the conclusion that the glass is relatively empty. In contrast, the unmarked frame might have been used either because the glass was previously emptier or simply because that linguistic form is generally more readily used.

The proposal advanced here would help explain the slight skew in the experimental data of McKenzie and Nelson (2003) and Sher and McKenzie (2006), which suggest a general tendency for speakers to use ‘full’ rather than ‘empty’ frames, and for hearers to derive more information from ‘empty’ frames than full frames. Nevertheless, as discussed above, the findings of Ingram et al. (2014) suggest that ‘full’ frames can convey reference point information, albeit more weakly, and therefore that reference point information can sometimes outweigh considerations of markedness.

8. Speaker Intentions

Before closing, I will consider an objection to the claim that reference point information is conversationally implicated, on the basis that it is not plausibly *intended* by the speaker. In this vein, Holleman and Pander Maat complain:

[T]he explanation put forward by Sher and McKenzie is not really communicative, as it does not include any reference to the communicators’ intentions. The central metaphor of leakage suggests an automatic process, beyond the control of the language user. (Holleman & Pander Maat, 2009, p. 2206)

As discussed in §2.2 it is not clear, in fact, whether implicatures must be thought of as part of speaker meaning at all. Here I want to suggest that, even if they are, this does not automatically disqualify reference point information from counting as implicated.

First, it is clear from the quote above that Holleman and Pander Maat make certain assumptions about the *nature* of speaker meaning: specifically, that it cannot be ‘automatic’ or ‘beyond the control of the language user’. However, I think this is far from obvious. Plausibly, at least some of the time, communicative intentions are relatively automatic and sub-conscious, with no need for competent speakers to engage in conscious deliberation about how to formulate an utterance. Of course, the question of precisely what kinds of psychological processes underpin communicative intentions is a large and complex one, which I cannot do justice to here.

However, it would at least require more work to show that speaker meaning has the properties assumed by Holleman and Pander Maat.

Second, I want to return to Grice's analysis of 'The king of France is bald', which led him to introduce the framing maxim. According to Grice, that statement implicates that hearers may only challenge the fact of the king's baldness, not his mere existence. Yet, it seems unclear whether this information is speaker-intended. At the very least, this case seems no *more* clear than that of reference point information. Certainly, it is rather implausible that a speaker of 'The king of France is bald' must have a *conscious* intention about the challengeability of the entailments. If that is correct, it suggests that reference point information could stand or fall along with other examples like this, which Grice took to be genuine implicatures. Either all of them should be considered to be speaker-meant (and therefore unproblematically implicated) or none should. And, in the eventuality that none should, it seems just as reasonable to drop the requirement that implicatures must be speaker-meant as it would be to weed out those of Grice's implicatures which are not speaker-meant (especially if we take seriously the exegetical observations made by Saul).

While I do not take these points to provide a knock-down argument, I suggest that they do, at least, push the burden back onto my opponent to provide stronger arguments against analysing reference point information as conversationally implicated. Given that an implicature-based analysis provides a parsimonious framework for capturing that information, I suggest treating it as such.

9. Conclusion

In this chapter, I have argued that the reference point information conveyed by frames fits the mould of a Manner-based conversational implicature. In the next chapter I will defend this proposal against competing analyses, according to which reference point information intrudes on what a frame 'says'.

CHAPTER 8

Keeping Semantics Minimal

Abstract

In the previous chapter, I showed how the reference point information conveyed by a frame could be analysed as a Gricean conversational implicature. In recent years, though, many have argued that information traditionally treated as implicated may intrude on what is ‘said’. There are different versions of this claim, which reanalyse implicatures as affecting truth-conditional semantic meaning, or the truth-conditions of what is asserted, or non-truth-conditional semantic meaning. I consider in turn how each approach might be extended to the case of reference point information. I argue that the resulting analyses lack plausibility or parsimony. I conclude that the framing data lend further support to a broadly Minimalist understanding of the semantics-pragmatics divide.

1. Introduction

Building on the discussion of the last two chapters, I will simply assume here that attribute frames convey reference point information, and that this explains framing effects at least in part. Thus, for example, frames (1) and (2) tend to convey the reference point information in (3) and (4), respectively:

(1) The glass is a quarter full.

(2) The glass is three quarters empty.

(3) The glass is relatively full.

(4) The glass is relatively empty.

Likewise, (5) and (6) tend to convey the reference point information in (7) and (8):

(5) This player made 40% of his shots last season.

(6) This player missed 60% of his shots last season.

(7) This player made a relatively large proportion of his shots last season.

(8) This player missed a relatively large proportion of his shots last season.

In general, the reference point hypothesis predicts that attribute frames, which include one or other of a pair of contradictory expressions, tend to convey that the subject of predication instantiates the corresponding property to a relatively high degree. As discussed previously,

this can explain why hearers' responses shift in response to alternative frames – in other words, why framing effects are observed.

In chapter 7, I showed how the reference point information conveyed by frames might be analysed as a Manner-based Gricean implicature. However, I will now consider an important challenge to that analysis. While Grice sought to hold implicatures apart from what a speaker 'says', a large post-Gricean literature has problematised his distinction. Many now argue that a subset of apparent implicatures actually intrude, in some sense, on what is 'said'. Mapping the arguments back to the spectrum of views surveyed in chapter 2, I will consider how each approach might be applied to the reference point information conveyed by frames.

The structure of the chapter is as follows: in §2 I consider Indexicalism-inspired approaches, according to which reference point information affects the truth-conditional semantic meaning of an uttered frame. In §§3-4 I turn to the idea that reference point information is a *free* pragmatic effect but one which still affects the truth-conditions of what is 'said'. I discuss Contextualist and Relativist versions of this claim. In §§5-7 I consider whether reference point information could be part of a frame's non-truth-conditional semantic meaning, exploring the role of presupposition, ambiguity, and 'bias'. I argue that the approaches discussed in §§2-7 are either implausible or less parsimonious than the implicature-based analysis proposed in the previous chapter. In §8 I argue that, insofar as explaining framing effects requires a distinction between a frame's minimal semantic meaning and the reference point information it implicates, this provides some further support for a form of Semantic Minimalism.

2. An Indexicalist Approach

As argued previously, since the reference point information conveyed by a frame is readily defeasible, it cannot arise *purely* as a result of composing the conventional meanings of the constituent words. Otherwise, it should arise in *all* contexts in which the frame is uttered by a competent speaker. Nevertheless, it remains possible that frames include certain context-sensitive elements, which contribute reference point information to their semantic contents. I will consider two versions of this Indexicalism-inspired analysis, the first of which posits overt context-sensitivity in the surface forms of frames, while the second appeals to covert variables in their underlying logical forms.

2.1. Overt Context-sensitivity

As discussed in chapter 2, some elements of language are clearly context-sensitive, including expressions like 'I', 'you', 'this', 'that', 'here', and 'today', the denotations of which vary across occasions of utterance. To recap, the context-sensitivity of these expressions is standardly

traced back to their conventional meanings, which demand contextual ‘saturation’.¹⁹¹ Recall that, standardly, context-sensitive expressions are thought to contribute occasion-specific denotations to the truth-conditional semantic meaning of a sentence. For example, ‘this’ will contribute the object to which it refers on each occasion of use. Therefore, if a frame were to contain a context-sensitive element, that element would contribute an occasion-specific denotation to its truth-conditional semantic meaning.¹⁹²

In fact, I think it is immediately implausible that reference point information arises from overtly context-sensitive elements of frames. The elements of surface structure that differ across alternative frames are the contradictory predicate expressions used (like ‘full’ and ‘empty’; ‘missed’ and ‘made’) and the complementary quantitative elements (like ‘a quarter’ and ‘three quarters’; ‘40%’ and ‘60%’). However, none of these plausibly requires saturation by a contextual reference point. Take, for example, the expression ‘missed’ in (6). In order to contribute reference point information, its denotation would have to be restricted to the set of individuals who miss *relatively large* proportions. In other words, to capture the reference point information, the conventional meaning of ‘miss’ would need to be something like ‘miss a relatively large proportion’. Importantly, this would apply across all occasions of its use (with occasion-specific factors simply determining the relevant reference point). It is clear from considering other syntactic contexts for ‘miss’ this cannot be its conventional meaning. For example, consider (9) below:

(9) The player missed his shot because he was distracted by the flashing cameras.

In (9) ‘miss’ does not plausibly mean ‘miss a relatively large proportion’. Indeed, I take it that there is no temptation to suppose that (9) appeals in any way to a reference point player who missed relatively few shots. Similar reasoning applies to the other expressions substituted across frames, suggesting that reference point information is not the result of straightforward overt context-sensitivity.

2.2. Covert Variables

Even though expressions like ‘full’, ‘empty’, ‘make’, and ‘miss’ are not plausibly context-sensitive, it remains possible that attribute frames contain context-sensitive elements in their logical form. To flesh out this idea, I will draw on the Indexicalist account proposed by Stanley

¹⁹¹ As before, I will remain neutral as to the degree to which conventional meanings constrain what context-sensitive expressions can denote, and also on whether saturation depends on speakers’ intentions or objective features of the situation.

¹⁹² Again, one might argue instead that semantics is exhausted by the composition of conventional meanings, as does Harris (forthcoming). And, as argued in chapter 5, I believe semantic meanings can be *truth-evaluable* on either approach. Here, though, I will assume the standard picture.

and others, which was introduced in chapter 2. Recall that these theorists appeal to covert variables, which contribute occasion-specific comparison classes, domains, or locations to truth-conditional semantic meanings. For example, the logical form of (10) can be represented roughly as in (11), where the variable 'x' requires saturation by an occasion-specific comparison class:

(10) John was old when he died.

(11) John was old(x) when he died.

Similarly, we might analyse the logical form of (6) as containing a placeholder for a contextually salient player or class of players, relative to which the target player is to be compared. We could represent this roughly as follows:

(12) The player missed(x) 60% of his shots.

This would make the semantic meaning of (6) dependent on whichever player (or set of players) is salient on the given occasion of use.

The problem with this analysis is that it does not explain why the player should be thought of as having missed a *relatively large* proportion of shots. It only requires that there must be *some* comparator in play. To capture the reference point information, we would need to place an additional restriction on the comparator, requiring it to render 60% a relatively large proportion of shots to miss. A Stanley-style account provides no motivation for imposing such a restriction. Moreover, once that restriction is imposed, it is unclear that the covert variable, x, is playing any substantive role in explaining how reference point information arises.¹⁹³

There are various other problems with this kind of approach. First, frames seem far less likely to generate intuitions of semantic incompleteness than, for example, sentences containing the expression 'old'. This limits the intuitive argument for positing a covert variable. Also, it may be difficult for the Indexicalist to explain why reference point information is defeasible, if it is supposed to arise from logical form. In light of these points (taken together with the more foundational criticisms of Indexicalism mentioned in chapter 2) I suggest that reference point information is unlikely to be traced back to either overt or covert context-sensitivity in the frame itself. Therefore, it does not plausibly affect a frame's truth-conditional semantics. In the next three sections, I will consider whether it could nevertheless affect what frames are used to 'say'.

¹⁹³ Hall (2008) makes a similar point in relation to a different set of cases.

3. Contextualism

As discussed in chapters 2 and 5, Contextualists and Relativists deny that Minimalists and Indexicalists carve at a natural joint when they focus on the lexical and syntactic features of sentences. Instead, it is argued that capturing truth-conditional contents with genuine theoretical value will require appeal to ‘free’ effects. I will begin by considering how a Contextualist account might be brought to bear on framing cases, turning to Relativism in §4.¹⁹⁴

As we saw in chapter 2, Contextualists are concerned with the truth-conditional meanings that are ‘said’ or ‘explicated’ (I will also talk about these meanings in terms of what speakers ‘assert’ or ‘directly communicate’).¹⁹⁵ The focus on what speakers assert tracks Grice’s requirement for what is ‘said’ to be speaker-meant. Contextualists depart from Grice, however, in weakening the link between ‘said’ contents and conventional meaning. In this section, I will consider whether reference point information could plausibly intrude on what is asserted. I do so by assessing it against the various criteria Contextualists have put forward for individuating explicatures or what is ‘said’. I will begin by focusing on Relevance Theory’s category of explicature, although in §3.4 I will briefly discuss Recanati’s distinct psychological criterion for what is ‘said’.

3.1. Formal and Functional Criteria

As we saw in chapter 2, explicatures are supposed to be ‘developments’ or ‘expansions’ of an uttered expression’s logical form. In other words, the information that is derivable via linguistic decoding should be contained within the explicature. As I have presented the reference point information so far, it would not meet this criterion: for example, (8) does not plausibly contain the logical form of (6) (both repeated below) since (8) includes no component corresponding to ‘60%’.

(6) This player missed 60% of his shots last season.

(8) This player missed a relatively large proportion of his shots last season.

However, we could easily reformulate the reference point information as follows:

(13) This player missed 60% of his shots last season, and thereby missed a relatively large proportion.

¹⁹⁴ As discussed in chapters 2 and 5, I will not address Occasion-sensitivity, which claims that communicated meanings cannot be adequately represented, either in language or thought, but is instead essentially context-involving. Whether or not this is correct, the claim is orthogonal to the question of whether reference point information is part of what is ‘said’ or implicated.

¹⁹⁵ As noted previously, Contextualists typically do not classify such meanings as ‘semantic’. Instead, that label is reserved for conventional aspects of meaning, which are held to be theoretically inert.

It seems possible that a hearer of (6) might form a representation like (13), which contains and expands the logical form of the frame.

Nevertheless, (13) arguably fails to meet the separate ‘locality constraint’ put forward by Carston and Hall (2012). This criterion requires that any modulation and enrichment operates on sub-propositional constituents of the logical form, rather than on a complete propositional meaning. In (13), the additional clause seems to operate on the proposition that *this player missed 60% of his shots last season*. Perhaps, then, we ought to represent the explicature as follows, where the added constituent specifically modifies the proposition:

(14) This player missed the relatively large proportion of 60% of his shots last season

If hearers of (6) formed the representation in (14), that could satisfy the first two formal tests for explicature. I believe (14) could also pass the ‘scope test’, since this content seems to survive the embedding of (6) in relevant syntactic environments. For example, consider the following sentence in which (6) is in the antecedent of a conditional:

(15) If this player missed 60% of his shots last season, then he won’t be selected for the team.

The antecedent of (15) still plausibly conveys (14). Indeed, the information in (14) provides the explanatory connection that links the antecedent with the consequent. As far as the formal diagnostics go, then, the reference point information conveyed by a frame like (6) could potentially be considered an explicature.

In addition to these formal tests, recall that Carston also proposes a *functional* constraint on explicatures, whereby they may act as premises for inferring implicatures but cannot themselves be entailed by implicatures. Plausibly, (14) could play the requisite role. Imagine, for example, that a speaker utters (6) in the context of selecting players for a basketball team. The hearer might arrive at something like the following conclusion:

(16) This player should not be selected for the team.

We can treat (16) as an implicature. It is inferable from (14), taken together with a further background premise that players who miss relatively large proportions of their shots should not be selected for the team. Meanwhile, (14) is clearly not entailed by (16): there are potentially many other reasons why a player should not be selected for the team, so it may not be inferred from the fact that *this* player should not be selected, that he missed a relatively large proportion of his shots. Thus, (14) seems capable of meeting Carston’s functional criterion.

We might be tempted, then, to conceptualise reference point information as intruding on what is ‘explicated’ by an uttered frame. The reference point information can bridge the gap between

a frame's logical form and the implicatures it is used to convey. However, these formal and functional tests should not be treated as conclusive. As acknowledged by Carston and Hall (2012), it is only to be expected that *formal* tests will ultimately be insufficient, since they operate at the level of sentence types, whereas explicatures are supposed to depend on token utterances (see also (Borg, 2016, p. 343)). Contextualists have also acknowledged that implicatures can sometimes meet the formal and functional criteria as well: for example, playing a similar intermediate role in the inference of further implicatures (Carston, 2002, p. 190).

It is more controversial whether implicatures can survive embedding under logical operators without their needing to be reclassified as explicatures. Borg (2009) is among those who argue that generalised conversational implicatures can do so, and that, therefore, the scope test fails to isolate explicatures. For example, Borg argues that the semantic truth-conditions of a conditional sentence are strictly independent of any information that might be implicated by its antecedent. She writes:

[T]he advocate of minimal semantics will have to bite the bullet with respect to the apparent semantic relevance of some implicatures, claiming that in all cases it is an *appearance* of truth-conditional relevance not a genuine case of rich pragmatic input to semantic content. Thus, given (3) and (4) (repeated here):

- 3) If Jill blew the whistle on poor practices at work and was sacked, then she is entitled to compensation.
- 4) If Jill was sacked and blew the whistle on poor practices at work, then she is entitled to compensation.

the minimalist must hold that there is no semantic difference between the two.

(Borg, 2009, p. 79, emphasis in original)

Borg accepts that the intuitive understanding of the antecedent in her conditional sentence (3) is that Jill was sacked *after* (and presumably *because*) she blew the whistle. In contrast, in her conditional sentence (4) the antecedent is intuitively understood as meaning that Jill was sacked first and *then* blew the whistle (which is why (4) sounds odd). Nevertheless, Borg maintains that our intuitions here are tracking *implicatures* conveyed by (3) and (4). The intuitions are particularly strong and compelling, she argues, because we derive generalised implicatures relatively mechanically and habitually. This does not show, however, that they are not implicatures.

Applying the idea to our framing case, neither (6) nor the conditional sentence (15) need include reference point information as part of an explicature; instead that information could

remain *implicated*. In other words, (15) (repeated below) would be semantically equivalent to (17):

(15) If this player missed 60% of his shots last season, then he won't be selected for the team.

(17) If this player made 40% of his shots last season, then he won't be selected for the team.

Both would have the same minimal semantic truth-conditions; it is just that their antecedents carry different implicatures: whereas the antecedent of (15) implicates that 60% is a relatively large proportion for the player to miss, the antecedent of (17) implicates that it is a relatively small proportion for the player to miss (or, equivalently, 40% is a relatively large proportion to *make*). Of course, (15) sounds far better than (17) but this can be explained by the fact that they carry distinct Manner-based implicatures. As discussed in the previous chapter, since reference point information depends on a regularity in use, it can be expected to arise in a fairly generalised fashion, coming along 'for free' with a frame unless defeated by competing considerations.

In sum, the formal and functional tests discussed above should be thought of as, at best, partial filters: while they may succeed in filtering out some implicatures, they will inevitably allow some others through. I believe this is what is happening with reference point information – it is an implicature that can sometimes meet the formal criteria which have been put forward for explicature. In the next subsection, I will argue that it is often blocked by more intuitive criteria.

3.2. Assertion

I want to suggest that, in many cases, reference point information will fail to count as something the speaker intuitively asserted. As noted in chapter 2, explicatures are supposed to track people's intuitive truth-value judgements, as probed by so-called 'context shifting arguments'. In many cases, I think it is unlikely that reference point information will impinge on these judgements. Consider, for example, the exchange below:

(18) A: What do you know about Federico?

B: He missed 60% of his shots last season.

Here B plausibly conveys the reference point information that Federico missed a relatively large proportion of his shots. However, had 60% in fact constituted a relatively small proportion of shots to have missed, it is rather *implausible* that B would have asserted something *false*. This suggests that the reference point information fails to affect the intuitive truth-conditions of B's assertion. In other words, it is not plausibly part of the speaker's explicature. Relatedly, I take it that a speaker like B intuitively incurs little or no liability for the false reference point

information. In other words, this information would fail to affect the speaker's 'conversational liability', which Borg (2017) proposes as the hallmark of explicature.¹⁹⁶

That said, we can construct contexts in which reference point information is more plausibly asserted. An example is the exchange considered in the previous chapter, and reproduced below:

(19) A: Why wasn't Federico recruited to the team?

B: He missed 60% of his shots last season.

As discussed previously, B's utterance can only be treated as a relevant reply to A's question if 60% is a relatively large proportion of shots to have missed. Perhaps in this case, then, B would be judged to have asserted something false, should it turn out that 60% in fact constitutes a relatively small proportion of shots to have missed. And, relatedly, we might intuitively hold B somewhat *liable* for communicating false reference point information in this context.

Although I want to grant this possibility, I believe it makes better sense to treat the reference point information as an implicature in *both* exchanges (18) and (19), with the only difference being in how *strongly* it is implicated. The proposal put forward in the previous chapter can capture both cases in a unified way: in the first case, where the reference point information is only supported by the framing maxim, the implicature is relatively weak. In the second case, where considerations of relevance reinforce the implicature, it is conveyed with greater strength. I submit that this univocal account provides a more parsimonious explanation of the data. Thus, even if we allow for a distinction to be drawn between cases in which reference point information affects intuitions about truth and liability and those in which it does not, we should consider that to be a difference in degree, not a difference in kind. As will be reiterated below, the key distinction for explaining framing effects remains the one between a frame's minimal semantic meaning (which renders it logically equivalent to its counterpart) and the reference point information it typically conveys (which renders alternative frames inequivalent to one another).

3.4. Primary vs. Secondary Processes

I end the section on Contextualism by turning briefly to the psychological criterion proposed by Recanati. As discussed in chapter 2, Recanati's criteria for individuating what a speaker 'says', partially overlap with those Relevance Theorists use for individuating explicatures. As we saw,

¹⁹⁶ I don't deny that B's utterance sounds odd where the reference point information is false, and might be considered misleading (as will be discussed in the next chapter). However, this can be adequately explained on an implicature-based approach.

though, Recanati emphasises a psychological criterion, whereby what is ‘said’ is the product of ‘primary’ pragmatic processes, operating unconsciously and sub-personally. In contrast, implicatures are thought to be the result of ‘secondary’ pragmatic processes that involve conscious, personal-level calculation.

Interestingly, Recanati’s picture would seem to get things exactly the wrong way around for framing cases: the contexts in which reference point information is most plausibly asserted (and therefore the ones in which it might affect what is ‘said’) are those where it is calculable on the basis of general rational considerations, such as relevance and informativity, with little need to appeal to a tacitly known regularity in how frames are used. Yet, on Recanati’s framework this would point to the reference point information being *implicated* rather than ‘said’. Conversely, the contexts in which there is less temptation to suppose that the reference point information is asserted are those in which we need to rely more heavily on tacitly known use properties, which presumably operate below the level of consciousness. On Recanati’s framework, however, these would seem to be classifiable as part of the *primary* processing machinery, which is supposed to deliver up ‘said’ contents rather than implicatures. Therefore, I do not think a Recanati-style Contextualism will help motivate the idea that reference point information intrudes on what is ‘said’.

4. Relativism

As discussed in chapter 2, Semantic Relativism is the view that sentences obtain different truth conditions on different occasions of utterance, not because they obtain different meanings but because different contexts require them to be evaluated for truth against parameters with different values. Taking inspiration from this approach, we might hypothesise that alternative frames have equivalent meanings but that their truth conditions can come apart in certain contexts. The idea would be that the reference point is a contextual parameter that affects truth-conditions without contributing to the meaning of an uttered frame. For example, if a player really did miss a relatively large proportion of shots, relative to the contextual reference point, then an utterance of (6) would be evaluable as true. If not, it would be evaluable as false.

I want to dispose of this possibility fairly swiftly. As discussed above, there is often little temptation to suppose that reference point information affects the intuitive truth or falsity of an uttered frame. To that extent, a Relativist approach seems unmotivated. As before, I am willing to grant that reference point information might affect intuitive truth-value judgements in *some* contexts, and the Relativist apparatus is potentially applicable to these cases. However, we would still be left with the task of providing a separate explanation for cases in which reference point information is conveyed *without* truth-conditions being affected. In contrast, the

implicature-based analysis proposed in the previous chapter brings both kinds of case under a single explanatory mechanism.

Before I move on to another set of approaches, I note that on both Contextualist and Relativist approaches, once the truth-conditions of an utterance are taken to depend on reference point information, there is supposed to be no theoretical role for a more minimal meaning to play. Against this, I think alternative frames continue to express logically equivalent semantic meanings *at the same time as* conveying distinct reference point information. Some phenomenological evidence for this is that one can appreciate the equivalence of two attribute frames like the ‘60% made’ frame in (5) and the ‘40% missed’ frame in (6) while continuing to feel that the first makes the player sound better than the second. Indeed, as discussed in chapter 1, this is what makes framing effects genuinely puzzling: it is not just that we form equivalent or inequivalent representations in response to frames but that we can feel pulled in both directions at the same time. I want to suggest that frames can sensibly and legitimately be considered equivalent (as experimenters have traditionally supposed), yet, in another sense they can be reasonably understood to be inequivalent. Therefore, whereas Tversky and Kahneman compare framing effects to visual illusions like the Müller-Lyer, a better analogy might be with the Necker cube, or the duck-rabbit: rather than our being misled by surface cues, there are two distinct sets of evidence operating simultaneously in different directions (which, in the framing case, correspond to semantic and pragmatic evidence). The phenomenology of framing effects thus provides some further *prima facie* evidence that minimal semantic meanings have a genuine psychological role to play alongside more enriched interpretations.

5. Presupposition

I turn now to consider a different set of theories, which have sought to enlarge semantics, accommodating linguistic phenomena that do not necessarily make a truth-conditional contribution.¹⁹⁷ I begin in this section by assessing whether reference point information could be *presupposed*. My focus here will be on presuppositions that are triggered linguistically.¹⁹⁸ For illustration, consider the sentence below:

(20) Federico regrets missing the shot.

¹⁹⁷ According to a still more radical view, semantic meaning might be thought of as *entirely* independent of truth-conditions (Chomsky, 2000; Pietroski, 2005, 2018). Thus, Pietroski (2015) seeks to explain framing effects by appealing to linguistic properties that are entirely detached from properties of the world. I will not pursue his proposal here, although it would be an interesting avenue for future research.

¹⁹⁸ On some theories, presupposition is taken to be a pragmatic phenomenon (Simons, 2006; Stalnaker, 1999). However, I set these aside since they would not involve ‘intrusion’ in the sense being discussed in this chapter.

In (20), the expression ‘regret’ is thought to trigger the following presupposition:

(21) Federico missed the shot.

This is not part of what is asserted when (20) is uttered but is rather a precondition for felicitously uttering (20).

Perhaps, similarly, attribute frames might trigger reference point information.¹⁹⁹ For example, perhaps (6) could presuppose (8) (both reproduced below):

(6) This player missed 60% of his shots last season.

(8) This player missed a relatively large proportion of his shots last season.

There is something initially suggestive about a presuppositional analysis. It seems right that reference point information is often ‘backgrounded’ in a similar way to information that is presupposed. This point is further supported by the fact that reference point information may pass the ‘hey, wait a minute’ test proposed by Von Stechow (2004). For example, one might felicitously respond to an utterance of (6) with: ‘Hey, wait a minute, 60% is a relatively small proportion to have missed’.²⁰⁰

Further investigation shows a presuppositional analysis to be implausible, however. Most importantly, the reference point information fails to exhibit the projection behaviour characteristic of presuppositions. As discussed by Beaver and Geurts (2014), presuppositions typically persist when the asserted content is embedded in various syntactic environments, especially under negation, and also in questions, possibility modals, and conditionals.²⁰¹ For example, the presupposed content in (21) – that Federico missed the shot – continues to project from the following sentences:

(22) Federico doesn’t regret missing the shot.

(23) Does Federico regret missing the shot?

(24) Maybe Federico regrets missing the shot.

¹⁹⁹ Note that whereas the presupposition in (20) is triggered by the simple expression ‘regret’, in the framing cases the presupposition would probably have to attach to a more complex construction (like ‘missed [some proportion of] his shots’). I will simply grant for now that such an analysis would be possible in principle. However, if it weren’t, then that would merely expedite my conclusion that reference point information is not presupposed.

²⁰⁰ Or, perhaps, the ‘cautious assent’ test described by Predelli (2013, p. 71) might be better in many contexts. This uses ‘Yes, but...’ as in: ‘Yes, but 60% is a relatively small proportion to have missed’. Predelli treats this as a diagnostic for ‘bias’, which I will consider in §7.

²⁰¹ As Beaver and Geurts note, though, there is still disagreement about the diagnostic value of each test.

(25) If Federico regrets missing the shot, he'll be motivated to keep practicing.

However, reference point information seems not to project from the same environments (or at least not as clearly). It strikes me as less plausible, for example, that (8) survives the embeddings of (6) in the following sentences:

(26) This player didn't miss 60% of his shots last season.²⁰²

(27) Did this player miss 60% of his shots last season?

(28) Maybe this player missed 60% of his shots last season.

(29) If this player missed 60% of his shots last season, he'll be motivated to keep practicing.

Note that (29) is similar to the conditional in (15), discussed above. As before, the reference point information survives embedding in the antecedent. Crucially, though, it does not *project* from the whole conditional, as (21) does from (25). In other words, (29) taken as a whole does not entail that the player actually did miss a relatively large proportion of his shots, whereas (23) *does* entail that Federico actually did miss the shot. This subtle but important difference suggests that we are dealing with an implicature rather than a presupposition.²⁰³ Given the projection data discussed here, I do not think reference point information is plausibly presupposed.

The ease with which reference point information can be defeated provides further support for the same conclusion. Strawson (1950) argued that, whenever a presupposition turns out to be false, the asserted content will be neither true nor false. For example, if Federico had not missed the shot, (20) would simply lack a truth-value; the question of its truth or falsity simply doesn't arise. Others, following Russell (1957), have thought that something *false* would have been asserted in that scenario. On either analysis, though, the speaker has certainly not said something *true*.

In contrast, as argued above, it seems possible for attribute frames to express truths, even when the reference point information turns out to be false. Thus, on either the original Strawsonian view, or the Russellian alternative, reference point information is not plausibly presupposed by attribute frames. Although Strawson himself later adopted a more moderate view of

²⁰² Negating the alternative, positively-worded, frame encourages an inference to the antonym, as discussed by Krifka (2007) and Ruytenbeek, Verheyen, and Spector (2017): in 'This player did not make 40% of his shots last season', the negated verb phrase 'did not make' is naturally interpreted as 'missed', yielding 'This player missed 40% of his shots last season'. The asymmetry between frames here may be linked to the phenomenon of markedness, discussed in the previous chapter.

²⁰³ A similar point is discussed by Davis (2016, §4.6). He gives the example of the sentence 'If they had a child and got married, the Church would disapprove', in which the antecedent continues to implicate that they had a child and *then* got married but this information does not project from the whole conditional.

presuppositions, presupposed information is still commonly understood to be *difficult* to cancel. For example, Beaver and Geurts (2014, §3) observe that cancelling a presupposition (unlike cancelling an implicature) is usually only possible when the trigger is embedded under an operator. Thus, one might sensibly utter (30), in which (20) is embedded under negation:

(30) Federico doesn't regret missing the shot because he *didn't* miss it.

In contrast, it is far harder to come up with a context in which it would be coherent to utter (31), in which (20) is unembedded:

(31) Federico regrets missing the shot but he didn't miss it.

Reference point information, on the other hand, is defeasible even in unembedded contexts, as has been discussed in previous chapters. Again, then, it seems to pattern more like an implicature than a presupposition. I believe that is how it should be classified. Before proceeding, though, I will briefly consider a 'presupposition denial' account of reference point information that has been put forward in the psycholinguistic literature.

5.1. Presupposition Denial

Moxey, Sanford and others originally developed their 'presupposition denial' account to explain the effects of certain 'natural language quantifiers' (NLQs) like 'few' and 'a few' (Moxey, 2011; Moxey & Sanford, 1987; Sanford, Dawydiak, & Moxey, 2007; Sanford, Fay, Stewart, & Moxey, 2002). Despite those two expressions being quantitatively equivalent, the 'positive' NLQ, 'a few', has interestingly different effects than the 'negative' NLQ, 'few'.²⁰⁴ Specifically, each brings into focus distinct sets of entities. To illustrate, consider the following stimuli, taken from (Sanford et al., 2002, p. 130):

(32) A few of the fans went to the match. They enjoyed it greatly.

(33) Few of the fans went to the match. They watched it on TV at home instead.

The use of 'a few' in (32) brings into focus the set of fans who *did* go to the match (which the theorists dub the 'reference set'), and makes these fans available for anaphoric reference: in the continuation, 'they' is most naturally read as referring to the fans who went to the match. In contrast, the use of 'few' in (33) brings into focus the set of fans who *did not* go to the match (the 'complement set'), making them available as the referents of 'they'. As demonstrated by (34) and (35) below, it is difficult to get the opposite readings, such that 'they' refers to the complement set when 'a few' is used, or to the reference set when 'few' is used:

²⁰⁴ The labels 'positive' and 'negative' are being used here as *linguistic* rather than *evaluative* descriptors. Sanford et al. (2007) describe a series of tests for distinguishing positive and negative NLQs in this sense.

(34) A few of the fans went to the match. They watched it on TV at home instead.

(35) Few of the fans went to the match. They enjoyed it greatly.

More generally, it is found that the reference set is typically in focus after the use of a positive NLQ, whereas the complement set is in focus following a negative NLQ.

The focus effects of negative NLQs have been explained in terms of ‘presupposition denial’ (Ingram & Moxey, 2011; Moxey, 2011; Sanford et al., 2007). The idea is that these expressions typically evoke a contrast with some *larger* quantity. Taking the example above, saying that ‘few’ of the fans went to the match presupposes an expectation that more fans would have attended. By establishing that expectation and simultaneously denying that it actually came to pass, the use of ‘few’ induces a focus on the resulting shortfall, making the complement set most readily available for anaphoric reference.

For the purposes of the current discussion, I am interested in the explicit connections the proponents of this account have drawn with the reference point hypothesis (Ingram & Moxey, 2011; Moxey, 2011). They write:

McKenzie and Nelson (2003) found that if a participant was told a glass was half full, s/he was more likely to infer that the glass had previously been empty. If told the glass was half empty s/he was more likely to say it had previously been full. These authors argue that this is due to a tendency to emphasise the variable that has increased relative to an implicit reference point... In our experiments the complement set represents the group that has increased relative to the reference point. The manner in which the information in our sentences is framed allows readers to draw inferences about the situation just as McKenzie and colleagues found with logically equivalent frames. In their experiments, the frame itself gives more information than simply the volume in the glass. (Ingram & Moxey, 2011, p. 395)

The authors do not fully develop the application of the presupposition denial account to framing cases. However, one way of doing so would be to hypothesise that a frame like (6) presupposes the existence of a reference point player who missed less than 60% of their shots. At the same time, it explicitly denies that the subject of predication – the *target* player – missed less than 60%, by asserting that he made 60%.

As argued above, I think it is ultimately implausible that frames *directly* presuppose reference point information. However, they may well do so *indirectly*. This is because, arguably, the reference point information itself presupposes the existence of a salient reference point. Take, for example, the reference point information in (8), repeated below:

(8) This player missed a relatively large proportion of his shots last season.

Plausibly, ‘relatively’ in (8) triggers the presupposition that there is some reference point, relative to which the player is being compared. Therefore, wherever a frame conveys reference point information, it might be understood as indirectly presupposing the existence of a contextual reference point. What I am suggesting, then, is that reference point information is an implicature, which itself carries a presupposition.²⁰⁵

Moreover, I suggest that hearers often take the existence of the presupposed reference point for granted, ‘accommodating’ it in the sense described by Lewis (1979). This is one way to explain how reference point information involves a kind of ‘backgrounding’. As suggested by the ‘leakage’ metaphor, there may be something indirect, implicit, and even illicit about the way frames convey information about contextual reference points. I will return to discuss these ideas more fully in the next chapter. For now, I simply want to reiterate that the reference point information is not *itself* a presupposition. I now turn to consider another way in which one might attempt to capture it within non-truth-conditional semantics.

6. Ambiguous Conventions

As discussed in chapter 2, Lepore and Stone (2015) posit a large class of semantic effects which have traditionally been classified as pragmatic but which, they argue, in fact depend on interlocutors’ knowledge of linguistic conventions. Let’s suppose (as mooted in the previous chapter) that reference point information is associated with frames via some kind of convention of language use. Taking inspiration from Lepore and Stone, then, we might hypothesise that frames are simply ambiguous between interpretations that do and do not include reference point information. The reference point information would thus no longer be thought of as a pragmatic enrichment but rather as one of the semantic meanings associated with the frame.

I think it would be a mistake to treat attribute frames as being straightforwardly ambiguous in this way. First, there is an important asymmetry between the two interpretations, which the ambiguity view fails to capture. Specifically, the interpretation that includes reference point information *entails* the one that doesn’t. In other words, the minimal semantic meaning of the frame persists whether or not reference point information is also conveyed. This makes the framing case importantly unlike standard cases of ambiguity, which enable one or other interpretation, but not both. The two readings of a frame are mutually consistent rather than mutually exclusive.

²⁰⁵ Geurts (2010) discusses the inverse scenario, where a presupposition conveys an implicature.

Second, recall that Lepore and Stone claim that all conventional effects get onto the ‘conversational scoreboard’, which is supposed to track information that is made publicly available to interlocutors during communication. As argued above, though, reference point information is typically not part of what speakers intuitively assert. If it gets onto the scoreboard at all in such cases, it must have an importantly different status from other information speakers communicate.

Lastly, even where reference point information plausibly *is* asserted, that should not obscure the role played by a more minimal semantic meaning. As discussed in chapter 2, Lepore and Stone may themselves retain a distinction between semantic conventions that affect truth-conditions and those that don’t. It is this distinction that is important for explaining framing effects. I believe the distinction separates conventions of *meaning* from conventions of *use*: meaning conventions guarantee that alternative frames share a set of truth-conditions, while the use properties of frames lead to divergent reference point information being conveyed. In line with the argument of the previous chapter, then, I suggest we treat the effects of language use (whether or not they are conventional) as *pragmatic* effects on interpretation; specifically, we can analyse them as conversational implicatures.

7. Bias

As we saw in chapter 2, Predelli (2013) argues that some expressions have conventional ‘biases’, which constrain the situations in which they may be used. Unlike the ‘characters’ of expressions, these biases have no effect on truth-conditions. To recap the earlier example, the expression ‘hurray’ can only be used in situations in which (roughly speaking) the agent has a favourable attitude towards the state of affairs being described. Conversely, ‘alas’ can only be used where the agent has an unfavourable attitude. However, these constraints, it is argued, do not affect the truth or falsity of the sentences in which they appear.

Applying Predelli’s framework to the framing case, let’s grant, again, that it is *conventional* for speakers to select frames in accordance with the existence of contextual reference points. In other words, it would be conventional for speakers to use (5) where the player *made* a relatively large proportion of his shots, and (6) where he *missed* a relatively large proportion. Correspondingly, competent language users, who are presumed to be sensitive to this convention, may infer reference point information from the speaker’s frame.

Note that the speaker’s frame could not *fully* constrain the context to include the associated reference point. This points to a disanalogy with uses of ‘hurray’ or ‘alas’. Whereas sincere uses of those expressions would seem to impose absolute constraints on the speaker’s attitude, the sincere use of an attribute frame clearly does not require the reference point information to

obtain. As we have seen, that information might be defeated altogether by conflicting considerations. Even where it is conveyed, I suggest that it may be probabilistic rather than deterministic: the speaker's choice of frame make it *more likely* that the context contains the associated reference point; but this is not guaranteed.

This is not necessarily an insurmountable problem, though, since something similar could be said of several other cases Predelli presents. For example, he discusses the co-extensional terms, 'belly button', 'navel', and 'umbilicus'. These certainly *tend* to be used when addressing different audiences (roughly, children, adults, and medical professionals, respectively). They therefore provide some information about the identity of the addressee. However, the evidence is defeasible and probabilistic in a similar way to reference point information. After all, it would still make sense (although it would clearly be less felicitous) to use the expression 'belly button' in addressing medical professionals. Perhaps, then, we should think of expressions' biases as *increasing the probability* of different contextual facts obtaining, rather than as necessarily *requiring* them to obtain. Once this is clarified, reference point information might be analysed as part of a frame's 'bias'.

The problem with weakening the definition of a bias in this way, however, is that it becomes less plausible that it counts as a genuinely semantic phenomenon, especially if we are thinking of semantic meaning as being constant and fixed rather than defeasible and probabilistic. Likewise, it becomes less clear what is gained by analysing reference point information as a (semantic) bias rather than as a (pragmatic) implicature. After all, implicatures already seem to be perfectly capable of carrying probabilistic, defeasible information about the context of utterance. Therefore, there is no obvious need to appeal to *bias* to capture those features of the reference point information conveyed by frames. It seems more parsimonious to bring the reference point information under Grice's general framework than to establish a separate category to do similar work.

8. The Semantics-Pragmatics Divide

To summarise the preceding discussion, I argued in §2 that the reference point information conveyed by attribute frames cannot be part of its truth-conditional semantic meaning. However, in §§3-4 I acknowledged that it might sometimes affect what a speaker intuitively asserts and, in §§5-7, I further acknowledged that reference point information could potentially be analysed as part of non-truth-conditional semantic meaning – specifically, as a 'bias'. I nevertheless suggested that, in both cases, the implicature-based analysis developed in the previous chapter would capture the relevant data more parsimoniously. In this section, I argue that the case study of framing provides some further support for locating the semantics-

pragmatics divide as Minimalists have proposed, between very minimal truth-conditional semantic meanings and broadly Gricean pragmatic meanings.

The question of where to locate the boundary between semantics and pragmatics might at first appear to be a purely terminological one, with different theorists being free to choose how they wish to apply those concepts. However, I believe this appearance obscures a substantive debate. After all, if our conceptions of semantics and pragmatics are to have genuine explanatory value, they must track real-world psychological or sociological phenomena. I suggest that the empirical case study of framing effects can be used to inform the conceptual distinction between semantics and pragmatics.

Over the course of the last three chapters, I have argued that the distinction we require to do justice to the ‘information leakage’ account of framing effects is one that contrasts the logically equivalent minimal semantic meanings of alternative frames with the distinct reference point information these frames pragmatically implicate. This fits well with how McKenzie and Nelson themselves position their account. As discussed in the previous chapter, they explicitly take the communication of reference point information to be *pragmatic*. The implicature-based analysis I have proposed shows how to cash out this claim. In discussing the equivalence of frames, meanwhile, the information leakage theorists rarely use the terminology of ‘semantics’.²⁰⁶ Nevertheless, it is evident from the following quotes that they are relying on something like literal sentence meaning:

[I]f reference points reliably influence speakers’ frame selection, then *frames carry information beyond their literal content*. Frames that are logically equivalent might, nonetheless, convey different information. (McKenzie & Nelson, 2003, p. 597, emphasis in original)

[L]ogical equivalence is well-defined (a pair of statements is logically equivalent if each member of the pair necessarily entails the other) and, provided some care is taken in translating between logical connectives and natural language connectives, straightforward to diagnose (Sher & McKenzie, 2006, pp. 468-469).

Therefore, I think the analysis they present coheres very well with a minimal conception of semantics. If we were to focus instead on the (Contextualist) distinction between what is *asserted* and what is *implicated*, disregarding minimal truth-conditional meanings, the reference point information could end up on different sides of the divide on different occasions, sometimes being part of what is asserted and sometimes being part of what is implicated. The

²⁰⁶ One exception is their description of the logical equivalence of frames as being determined by ‘semantic entailment’ (Sher & McKenzie, 2011, p. 50).

sense in which frames remain equivalent in *all* of these contexts would be lost. Alternatively, if we were to focus on the (Conventionalist) distinction between *conventional* and *non-conventional* effects on meaning then, assuming that reference point information arises from a linguistic convention, the distinction required for the information leakage account would be collapsed entirely: both the equivalence and inequivalence of frames would be explained at the conventional semantic level, with pragmatics doing no substantive explanatory work. On either approach, then, we would end up muddling the distinction between the minimal semantic equivalence of frames and the divergent reference point information they convey. This tells against an expansive, ‘intrusive’, analysis: the distinction we need for explaining framing effects is one that separates *semantic meaning* from *pragmatic use properties*. In sum, the information leakage account of framing effects rests on a distinction between minimal truth-conditional semantic meaning and broadly Gricean pragmatics. By revealing and developing the philosophical underpinnings of the account, I hope to have set it on a firmer foundation.

Equally, given that the explanation of framing effects depends on a broadly Minimalist conception of the semantics-pragmatics divide, this case study can be understood as providing one more piece of evidence in support of Semantic Minimalism. Since the Minimalist’s way of drawing the boundary also tracks key aspects of how ‘semantics’ and ‘pragmatics’ have traditionally been understood in the philosophy of language – whereby semantics is conventional, constant, and language-driven, while pragmatics operates freely and defeasibly, in accordance with general rational principles of communication – that seems like a good result.

On a concessive note, I recognise, of course, that there are many other applications for the concepts of semantics and pragmatics aside from explaining framing effects, and these might ultimately point to a different way of drawing the boundary. That said, I hope at least to have shown that the case study of framing is one which puts minimal semantic meanings to work in an explanatorily valuable way.

9. Conclusion

In conclusion, I want to draw together the discussion in chapters 7 and 8 by identifying two senses in which framing might be considered a ‘contextual’ phenomenon. First, frames constitute part of the context in which minimal semantic meanings are expressed – they are the words *in* which these meanings are ‘clothed’. Therefore, since reference point information depends on which particular frame is used, it can be understood as arising from context in an analogous way to other pragmatic implicatures.

A second sense in which framing effects are contextual is due to the relationship between frames and contextual reference points: since cooperative speakers tend to use frames in

accordance with such reference points, hearers may infer information *about* the wider context in which a frame is uttered. Thus, frames are not just part of the context in which a meaning is expressed but also tell us something about other features of that context. In the final chapter of the thesis, I turn my attention to some of the real-world contexts in which frames are used and, drawing on the preceding analysis, I consider how we should assess the behaviour of both speakers and hearers.

CHAPTER 9

Framing: The Fault Lines

Abstract

In this chapter I assess whether the information leakage account provides a *rationalising* explanation of framing effects. I show how, on the one hand, hearers can no longer be said to violate the rational choice principle of description invariance. On the other hand, though, it does not automatically follow that hearers are always rational to respond differently to alternative frames. I sketch how a rational hearer's interpretation will depend on a reasonable assessment of the speaker's sincerity, reliability, and relevance. Although ordinary hearers may often fall short of the rational ideal, I argue that speakers, too, are often to blame for hearers' misinterpretations. Using real-world case studies, I show how the theoretical framework developed in earlier chapters grounds a more nuanced analysis of framing effects than standardly emerges from the psychological literature.

1. Introduction

As discussed in the introduction to the thesis, psychologists have standardly taken framing effects to show that hearers' behaviour falls substantively and systematically short of the rational ideal. As Bermúdez puts it: "The consensus view is that susceptibility to framing effects is a paradigm of irrationality" (Bermúdez, 2018, p. 180).²⁰⁷ Against the received view, the proponents of the reference point hypothesis purport to provide a *rationalising* explanation of framing effects. They address the following claim made by Kahneman:

Framing effects are less significant for their contribution to psychology than for their importance in the real world...and for the challenge they raise to the foundations of a rational model of decision making (Kahneman, 2000, p. xv).

In response, they write:

²⁰⁷ It is perhaps unfair to attribute this strong view to Tversky and Kahneman, the instigators of framing research. Commenting on the results of their early empirical investigations, they write:

These observations do not imply that preference reversals, or other errors of choice or judgment...are necessarily irrational. Like other intellectual limitations discussed by Simon...under the heading of "bounded rationality," the practice of acting on the most readily available frame can sometimes be justified by reference to the mental effort required to explore alternative frames and avoid potential inconsistencies. (Tversky & Kahneman, 1981, p. 458)

Here they explicitly allow that it might be rational overall for humans to make systematic errors, given the cognitive and environmental limitations we face. Nevertheless, susceptibility to framing effects is still ultimately seen as a departure from how an unboundedly rational agent would act. As we will see, the reference point hypothesis challenges even this weaker claim by proposing that our normative theory of rationality is consistent with, and may actually require, different responses to alternative frames.

Attribute framing, at least, does not raise a “challenge...to the foundations of a rational model of decision making”. (Sher & McKenzie, 2006, p. 488)

The reason they give is that “attribute framing effects are naturally justifiable in terms of sound choice-relevant inferences from the speaker’s choice of frame” (ibid). In this chapter, I will endorse a qualified version of this claim.

The structure of the chapter is as follows: in §2 I re-examine why framing effects are standardly thought to violate the rational principle of description invariance, and I show how the reference point hypothesis undermines the argument. In §3 I observe that it would be irrational for hearers to infer reference point information if it has been (explicitly or implicitly) defeated. In §4, I argue that the information may only be rationally inferred if the speaker is presumed to be adhering to Grice’s maxim of Quality. In §5, I argue that the speaker must also be presumed to have reliable beliefs. In §6, I argue that the salient reference point must be presumed relevant. Finally, in §7, I show how these arguments extend beyond attribute framing, to risky-choice paradigms. Throughout the discussion, I emphasise the responsibilities borne by *speakers* to communicate true, evidenced, reliable, and relevant information. To the extent that speakers violate communicative norms in conveying false reference point information, I conclude that they must attract some fault as epistemic agents.

2. Description Invariance

As discussed in chapter 1, it is standardly deemed irrational for hearers to respond differently to alternative frames. Specifically, this is held to violate the principle of description invariance. Spelling out this principle precisely is a difficult task.²⁰⁸ Below is a first stab:

Invariance Principle 1 (IP1): if two linguistic stimuli represent the same situation, they should give rise to the same judgements.

This lines up with what Tversky & Kahneman say in the following passage:

An essential condition for a theory of choice that claims normative status is the principle of invariance: different representations of the same choice problem should yield the same preference. That is, the preference between options should be independent of their description. (Tversky & Kahneman, 1986, p. 253)

However, IP1 is clearly implausible for agents with only limited knowledge: we may simply fail to realise that the two frames represent the same situation, and this need not be evidence of

²⁰⁸ For further discussion, see (Bermúdez, 2009, 2018).

irrationality but simply ignorance.²⁰⁹ In fact, Tversky & Kahneman themselves recognise that mere sameness of situation does not rationally require sameness of judgement. The quote from their 1986 paper continues:

Two characterizations that the decision maker, on reflection, would view as alternative descriptions of the same problem should lead to the same choice – even without the benefit of such reflection. (Tversky & Kahneman, 1986, p. 253)

This points to a different formulation of the principle, as follows:

IP2: If two linguistic stimuli would, on reflection, be viewed as representing the same situation, they should give rise to the same judgements.

However, IP2 raises other problems. First, it is unclear what would count as a sufficient amount of reflection. Second, there is no obvious way to operationalise the criterion: what is the right way to reveal someone's reflective view of what a linguistic stimulus represents?²¹⁰ In response to both questions, Kahneman and Tversky seem to have in mind whatever opinion an experimental participant would report, when presented with both frames. Thus, they write:

Invariance requires that the preference order between prospects should not depend on the manner in which they are described. In particular, two versions of a choice problem that are recognized to be equivalent when shown together should elicit the same preference even when shown separately. (Kahneman & Tversky, 1984, p. 343)

In a similar vein, Kahneman says:

It is the decision maker who should determine, after due consideration of both problems, whether the differences between them are sufficiently consequential to justify different choices. Violations of this lenient form of invariance demonstrate incoherence without a need for any judgement from on high about what is truly equivalent. (Kahneman, 2000, p. xv)

Perhaps, then, we might reformulate the invariance principle as follows:

²⁰⁹ Similar points are made by Sher and McKenzie (2006, p. 469, footnote 2) and Bermúdez (2018, p. 183).

²¹⁰ Sher and McKenzie also point out that it is rather unclear what we should conclude from explicit judgements about equivalence anyway:

[T]wo descriptions of a decision problem are said to be “equivalent” if, on reflection, people would endorse their equivalence. However, no normative theory of decision making has anything to say about the correctness of people's beliefs about equivalence – which, as we argue here with reference to most framing researchers' beliefs, may after all be wrong. (Sher & McKenzie, 2006, p. 469, footnote 2)

IP3: If two linguistic stimuli would be reported to represent the same situation when the stimuli are presented together, they should yield the same judgements when presented separately.

The problem with *IP3* is that presenting two linguistic stimuli together constitutes a qualitatively different speech act from that of presenting just one. Therefore, it would not be surprising if different information were pragmatically conveyed in each case. This amounts to a rejection of the claim made in the 1984 quote from Kahneman and Tversky: *we can't* assume that the same response should be elicited by alternative frames under joint and separate presentation conditions. It also suggests that the data from within-subjects framing studies, discussed in chapter 1, provides inconclusive evidence that experimental participants believe alternative frames to be equivalent *simpliciter*.

In fact, the reference point hypothesis gives us some reason to predict that joint and separate presentation contexts *should* elicit different responses. For example, take sentences (1) and (2) again, which tend to convey (3) and (4), respectively:

(1) This player made 40% of his shots last season.

(2) This player missed 60% of his shots last season.

(3) This player made a relatively large proportion of his shots last season.

(4) This player missed a relatively large proportion of his shots last season.

The meanings of (3) and (4) are contradictory. Therefore, if hearers are presented with both frames (1) and (2) together, they cannot consistently infer both associated pieces of reference point information: at least one will need to be discarded (and perhaps both).²¹¹ In contrast, under separate presentation conditions, the associated reference point information can be straightforwardly retained.

Moreover, assuming that the integration of reference point information may happen below the level of consciousness, it is not surprising that hearers typically fail to *report* that alternative frames convey distinct reference point information when presented separately (even if that is what's happening).

²¹¹ Thierman (2014) investigates joint presentation contexts (i.e. 'double framing'). Her results suggest that, in these contexts, responses pattern similarly to when only the positive frame is presented. There are different ways this could be explained: for example, it might be that the unmarkedness of the positive frames used in Thierman's experiments blocks the inference of reference point information. What is important for now is that joint and separate presentation contexts should not automatically be expected to elicit the same responses.

Once we have the information leakage account on the table, then, we can see how IP3 directly begs the question against it. IP3 assumes that the only relevant information frames can convey is their context-invariant meaning (which remains constant across joint and separate presentation conditions); this is precisely what the information leakage theorists deny.

Further, it is unclear why we should privilege joint presentation contexts over separate presentation contexts. Tversky and Kahneman (1986) describe joint presentation contexts as ‘transparent’, and separate presentation contexts as ‘non-transparent’. However, these labels simply assume, without justifying, the ability of the former to provide better access to the truth. Admittedly, there are sometimes good reasons for presenting multiple stimuli to participants, rather than just one. Hansen (2014) surveys psychological studies that directly compare judgements in joint and separate presentation contexts. He argues, with Hsee, Blount, Loewenstein, and Bazerman (1999), that some attributes are difficult to evaluate independently of relevant contrastive information. For example, the price one is willing to pay for 7oz of ice cream in a 5oz cup (where the amount of ice cream seems generous relative to the size of the cup) may fall relative to the price one is willing to pay for 8oz of ice cream in a 10oz cup (where the amount of ice cream seems ungenerous relative to the size of the cup) once one is able to evaluate both options together. I agree that contrastive information is often necessary for evaluation; indeed I will reiterate this point in §6. However, notice that joint presentation of alternative frames provides no additional contrastive information, unlike the joint presentation of the ice creams. Instead, the *same* attribute is predicated of the *same* player using different words. Indeed, joint presentation of alternative frames may even end up providing *less* contrastive information than presenting each separately, given that the reference point information associated with each frame becomes mutually undermining. Therefore, it is unclear why we should expect joint presentation contexts to be more reliable here.

Likewise, it is unclear why the contexts Kahneman and Tversky describe should be privileged over various conceivable others. For example, if participants were invited to reflect on the reference point hypothesis, perhaps they would then judge the frames to represent different situations. What’s to say that the basic joint presentation context elicits their considered views more reliably than this one?

I suggest, then, that IP3 is not a plausible characterisation of the invariance principle: as with the previous iterations, the antecedent of the conditional can obtain independently of the consequent. Instead, I propose that we appeal to something along the lines of Frege’s ‘Intuitive Criterion of Difference’, as discussed in chapter 3. To recap, this is characterised by Evans as follows:

[T]he thought associated with one sentence, *S* as its sense must be different from the thought associated with another sentence *S'* as *its* sense, if it is possible for someone to understand both sentences at a given time while coherently taking different attitudes towards them, i.e. accepting (rejecting) one while rejecting (accepting), or being agnostic about, the other. (Evans, 1982, pp. 18-19)

Although this is proposed as a criterion for individuating Fregean senses, the core idea can be adapted for the purposes of characterising the principle of description invariance. In particular, we can think of two linguistic stimuli as carrying different information if it is possible to understand both while coherently responding differently to each. This is reflected in IP4 below:

IP4: If it is impossible to understand what two linguistic stimuli mean without understanding them to represent the same situation, then they should yield the same judgements.

Crucially, if the information leakage account is correct, then alternative frames will not meet the condition in the antecedent of IP4: it *is* possible to understand what they mean while also taking them to represent different situations – specifically, situations in which contextual reference points take distinct values.²¹² The information leakage account thus takes an important first step in providing a rationalising explanation of framing effects. It suggests that hearers do not fall foul of the rational principle of description invariance when they respond differently to alternative frames. Instead, they may be displaying a rather sophisticated sensitivity to the pragmatic properties of utterances. Keren puts the point nicely:

Language conveys both explicit (semantic) and implicit (pragmatic) information. The fact that participants in experiments are able to decipher and infer some of the subtle implicit information contained in a message should perhaps not be used as evidence of failure of rational reasoning. The alternative is to incorporate participants' abilities in the description of rationality. One may even argue that failing to use the subtle hidden cues concealed in different frames may, by itself, be considered irrational (in the broader sense of the term). Thus, contrary to a recent suggestion by Kahneman (2000) that “the ubiquity of framing effects demonstrates that the human mind is not assigned to achieve coherence” (p. xv), in the alternative view framing effects, rather than exhibiting inconsistency, constitute a demonstration of a sophisticated use of language that normative accounts fail to take into account. (Keren, 2011, pp. 26-27)

²¹² Note that this remains consistent with the argument presented in chapter 3. As argued there, the distinct interpretations of alternative frames cannot be traced back to *conventional* meaning. Instead, it is due to the Manner-based implicatures they typically carry, as discussed in chapter 7.

It still remains to be seen, though, whether hearers behave rationally, all things considered, in responding differently to alternative frames.²¹³ Answering this question will be the focus of the remainder of the chapter. Since I will be interested primarily in whether hearers can justifiably infer reference point information, I will mainly focus on *epistemic rationality*, which concerns the truth-conduciveness of an individual's belief-formation process (how likely it is that the beliefs they form will track the way the world actually is). I will be less interested, for example, in *practical or ecological rationality*, which concern an individual's ability to survive and thrive in their environment (although presumably these notions are related).²¹⁴

3. Defeaters

The first point to note is that a hearer would be epistemically irrational to infer reference point information from a speaker's use of frame on occasions where the information is explicitly or contextually defeated (as discussed in chapter 7). Provided that there are no overriding factors, however, hearers are at least justified in attributing to the speaker an *intention to communicate* the reference point information.²¹⁵

This way of putting things is a little too simplistic, of course, since it assumes that hearers' inferences must be an all or nothing affair: either they attribute a particular mental state to the speaker or they don't. A more nuanced account would assign hearers' attributions of mental states some degree of probability. Strictly speaking, then, the claim I want to make at this stage is that, in the absence of defeaters, the speaker's use of an attribute frame should increase the probability which the hearer assigns to her intending to communicate reference point information. Bearing in mind the discussion of markedness in chapter 7, the use of a *marked* frame should increase this probability by more than the use of an unmarked frame. For ease of exposition, in what follows I will tend to describe hearers' inferences as if they were absolute. However, this should be understood as shorthand for their assigning greater probability to certain states of affairs.

²¹³ Therefore, I think McKenzie and Sher are a little quick when they claim:

For framing effects to raise normative concerns, they must violate a revised principle of description invariance, which states that *information equivalent* descriptions must lead to identical decisions. (McKenzie & Sher, 2008, p. 84, emphasis in original).

I will argue that various other normative concerns need to be dealt with before framing effects can be accommodated within a general theory of rational behaviour.

²¹⁴ Bolinger (2017) similarly invokes the notion of epistemic rationality in discussing when hearers are justified in taking offense from a speaker's use of a slur. She argues that they may be so justified, even if no offense is intended by the speaker, insofar as the term is commonly associated with derogatory attitudes. I will make a similar claim in §4 with respect to frames, arguing that hearers can be rational to infer reference point information, even where the speaker did not intend to communicate it.

²¹⁵ In the criteria for communicative success listed by Harris (2019), this equates to 'uptake': hearers recognise that the speaker uttered the frame intending for them to accept the reference point information.

4. Communicative Quality

A second condition for the rational inference of reference point information is that the hearer must reasonably presume the speaker to be adhering to something like the Gricean maxims of Quality, reproduced below:²¹⁶

1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

It would be irrational for hearers to attribute to the speaker a belief in the reference point information if they have reasons to suspect the speaker is violating the Quality norms. Instead, then, rational inference of reference point information depends on having a reasonable expectation that the speaker believes the information to be true, and that she has sufficient evidence for it. The notion of a 'reasonable' expectation here remains underspecified, of course, and there are undoubtedly debates to be had about precisely when a speaker can reasonably be expected to be adhering to Quality norms. I will assume that, while the condition may be met in many ordinary cooperative exchanges, it is possible to point to clear cases in which it is unfulfilled. In the next two subsections, I will consider two such cases: first, where the speaker has beliefs that are incompatible with the reference point information, and second where she is simply indifferent to its truth or falsity.

4.1. Disbelief

Imagine that a speaker utters (2) intending to communicate the reference point information captured by (4) (both repeated below) while simultaneously believing that the player in fact missed a relatively small proportion of his shots, compared with the salient reference point.

(2) This player missed 40% of his shots of his shots last season.

(4) This player missed a relatively large proportion of his shots last season.

If it is predictable that the speaker will attempt to convey false reference point information (say, because she dislikes the player's team or is known to have provided misleading testimony about

²¹⁶ Strictly speaking, these maxims apply to what speakers 'say'. However, I will assume that they generalise to other contents communicated by speakers, including what they implicate. I believe this remains in the spirit of Grice, who writes:

I would like to be able to show that observance of the Cooperative Principle and maxims is reasonable (rational) along the following lines: that anyone who cares about the goals that are central to conversation/ communication (such as giving and receiving information, influencing and being influenced by others) must be expected to have an interest, given suitable circumstances, in participation in talk exchanges that will be profitable only on the assumption that they are conducted in general accordance with the Cooperative Principle and the maxims. (Grice, 1989, pp. 29-30)

other players) then it would be epistemically irrational for hearers to attribute to her a belief in (4). All consequent inferences should then similarly be blocked; so, for example, the reference point information in (4) should not be inferred to be *true* on the basis of the speaker's utterance, nor used in making evaluative judgements.²¹⁷

In any scenario like this, where the speaker conveys reference point information insincerely, she is at fault: by failing to respect the first maxim of Quality, she violates an important (perhaps the *most* important) communicative norm. Crucially, the fault incurred by the speaker is unaffected by whether or not hearers can reasonably *expect* her to implicate false information. In other words, regardless of whether hearers behave in epistemically reckless ways, causally contributing to their being misled, that does not in any way mitigate the speaker's responsibility for *attempting* to mislead.

This last point is inspired by Saul's observation that "being partly causally responsible for a wrong done to one does nothing to alter the nature of that wrong" (Saul, 2013, p. 83). To illustrate, she provides the hypothetical example of a 'reckless' victim, who is mugged upon entering a part of town he knows to be dangerous. Even though the reckless victim is partly causally responsible for being mugged, this does not absolve the perpetrator from fault to any degree: his misdeed is equally as bad as that of mugging a 'careful' victim in a generally safe part of town. By analogy, where a speaker is predictably insincere, the responsibilities of speaker and hearer remain independent of one another. Any fault attributed to the hearer is additional to that of the speaker. In other words, the overall amount of fault increases, rather than a fixed quantity of fault being redistributed.²¹⁸

In the case of deliberate misleading considered here, it was stipulated that the speaker did, in fact, intend to communicate reference point information while believing it to be false. However, in real-world cases, intentions are less transparent and are often difficult to assess. Consider, as an example, the following advert for the 'Natural Cycles' contraceptive app, which appeared on various social media in August 2019:

²¹⁷ Notice how the case of misleading differs from the earlier case where the reference point information was defeated: the misleading speaker intends to convey the reference point information, and does nothing to block it, whereas in defeat case, the speaker does not intend to convey the information and takes it to have been (implicitly or explicitly) defeated. Of course, she might be wrong about its *actually* having been defeated, leading to inadvertent miscommunication, but this kind of scenario remains conceptually distinct from the case of deliberately misleading.

²¹⁸ Of course, the wrongs of misleading and mugging differ in nature and severity, and I do not mean to suggest that they have the same moral implications. Indeed, it is beyond the scope of this paper to consider how we ought to sanction, regulate, or generally respond to cases of insincerity (or unreliability or irrelevance). My aim here is primarily to describe the various ways in which frames can convey false information.

Introducing the first-ever contraceptive app! Natural Cycles is 93% effective with typical use and 98% with perfect use, it's completely hormone-free, CE Marked, cleared by the FDA, and perfect for anyone looking for a natural and healthy alternative ==> <https://www.naturalcycles.com/en>

Natural Cycles learns your unique cycle and identifies ovulation by analysing your basal body temperature which you measure when you wake up in the morning.

If you're looking for a natural contraception option to give your body a break from hormones, download the app created by female particle physicist, Dr. Elina Berglund and her husband Dr. Raoul Scherwitzl. Learn more about Natural Cycles and get started today ==> <https://www.naturalcycles.com/en>

**CE Marked,
Digital Contraception**

25K+ Reviews
★★★★★
4.8 out of 5

The advert makes the following claim:

(5) Natural Cycles is 93% effective with typical use and 98% with perfect use.

If the reference point hypothesis is correct, we can expect hearers to infer from the use of (5) that the sellers believe the following:

(6) Natural Cycles is relatively effective with typical use

(7) Natural Cycles is relatively effective with perfect use.

In particular, hearers may infer that the sellers believe the following, since the product is marketed as an alternative to the oral contraceptive pill:²¹⁹

(8) Natural Cycles is more effective than the pill with typical use.

(9) Natural Cycles is more effective than the pill with perfect use.

²¹⁹ That this is the intended reference point is apparent from the description of the app as 'hormone-free', which immediately renders the pill the salient alternative. The pill is also the most common form of contraception women receive from doctors and pharmacies in Great Britain: according to data obtained by The Guardian under the Freedom of Information Act, almost nine in 10 women who received contraception from the GP or pharmacies in 2017/18 were taking either the combined pill or "mini pill": <https://www.theguardian.com/uk-news/2019/mar/07/revealed-pill-still-most-popular-prescribed-contraceptive-in-england> (accessed on 20th August 2019).

According to official published statistics, the contraceptive pill is 91% effective with typical use.²²⁰ This makes (8) true.²²¹ However, the pill's effectiveness with perfect use is *at least 99%*. This makes (9) false.

Presumably, the sellers are aware that (9) is false, since they are marketing the app as a direct competitor to the pill. Therefore, it would have been less misleading for them to frame the information in terms of the proportion of failures, as in (10), in order to convey the accurate reference point information in (11):

(10) Natural Cycles fails in 2% of cases with perfect use.

(11) Natural Cycles fails in a larger proportion of cases than the pill with perfect use.

This is not necessarily a case of *deliberate* misleading. It is possible that the advertisers had no intention to convey (9). Perhaps, instead, they chose the 'effectiveness' framing primarily to formulate the information about typical and perfect use in a simple and elegant way. It does not follow, of course, that the sellers are absolved of all responsibility for hearers inferring false reference point information. Arguably, they ought to have taken more care to block that information, particularly given the high stakes involved.²²² Whatever the correct analysis of this particular case, then, the general point is that speakers may still be at fault, even when they mislead inadvertently, if it is deemed that they should have been more careful with their framing choices. One advantage of bringing to light the information 'leaked' by frames is that the scope for speakers to escape liability for certain misleading communications may, in time, be reduced. After all, once it is widely understood that frames tend to implicate reference point information, the use of an inappropriate frame cannot be attributed to mere ignorance.

4.2. Presupposition Failure

A different way of using frames misleadingly involves presupposition failure. As discussed in the previous chapter, the reference point information conveyed by a frame arguably presupposes the existence of a contextual reference point. For example, the use of 'relatively' in (3) or (4) presupposes the existence of a reference point relative to which the player made or missed a relatively large proportion of his shots. Accordingly, one way for a speaker of (1) or (2)

²²⁰ UK National Health Service statistics: <https://www.nhs.uk/conditions/contraception/how-effective-contraception/> (accessed on 26th September 2019).

²²¹ I am assuming that the Natural Cycles statistics are accurate. An investigation by the Swedish Medical Products Agency found that the rate of reported unintended pregnancies among Natural Cycles users between January and June 2018 was consistent with the 93% figure (full report available at: <https://lakemedelsverket.se/english/All-news/NYHETER-2018/The-assessment-of-Natural-Cycles-is-completed/>).

²²² The following article contains a personal account of the impact of the Natural Cycles app failing: <https://www.theguardian.com/society/2018/jul/21/colossally-naive-backlash-birth-control-app>.

to mislead is to believe that there is *no* contextually salient reference point. In such a scenario, it is not that the reference point information is straightforwardly believed to be false (as in the case considered above) but rather it is believed to be infelicitous because of the absence of a salient reference point.

As before, whether or not *hearers* are justified in attributing to the speaker a belief in the existence of a contextually salient reference point will depend on the extent to which the speaker can reasonably be expected to mislead. In many cases, even where no reference point is salient to the hearer, I suspect that there will be ‘accommodation’, in the sense discussed by Lewis (1979). In other words, the hearer will assume that there is *some* reference point in the context, which the speaker considers to be sufficiently salient.

From the speaker’s perspective, some cases of presupposition failure may be deliberate. To take a recent example, in August 2017 the British newspaper, The Telegraph, ran the following headline:²²³

(12) One in five unemployed people in the UK are migrants, official figures reveal for the first time.

According to the reference point hypothesis, hearers may infer that the writer intends to communicate the following:

(13) A relatively large proportion of unemployed people in the UK are migrants.

The implicature in (13) does not appear to have been defeated. On the contrary, it is further reinforced by the overall content and phrasing of the article, as well as background knowledge of the paper’s stance on immigration.

It is less clear, though, whether the headline writer should be attributed a *belief* in (13), partly because it is unclear what the contextually salient reference point is in this case. As is explicitly stated, this was the first release of official figures on migrant unemployment; therefore, the reference point cannot be a *prior* figure. The headline makes two other comparisons salient: first, the proportion of *employed* people in the UK who are migrants (as opposed to the proportion of *unemployed* people).²²⁴ Second, the headline may evoke a contrast with the proportion of unemployed people who are migrants *in other countries* (as contrasted with the

²²³ The full article is available at the following link:

<https://www.telegraph.co.uk/news/2017/08/16/one-five-unemployed-people-uk-migrants-official-figures-reveal/>. Similar headlines appeared in other newspapers, including The Sun and The Daily Mail.

²²⁴ Strictly speaking, the contrast should be with non-unemployed people rather than employed people, since some people count neither as employed nor unemployed, say if they are retired or in full time education. However, I will ignore this subtlety for now.

proportion *in the UK*). However, the full article discusses neither of these comparisons. Instead, the focus is on the general rise in UK immigration. Plausibly, then, the headline-writer may have believed that there was no contextually salient reference point and was therefore being deliberately misleading. Alternatively, s/he might merely have been indifferent to the existence of a reference point, as will be discussed in the next section. In other cases, presupposition failure may be inadvertent, if the speaker did not intend to convey the reference point information associated with their frame. However, this seems rather unlikely here.

In cases like the Telegraph example, there are potentially high stakes associated with the communication of reference point information. The idea that relatively many unemployed people in the UK are migrants stokes anti-immigrant sentiment, which may lead to increases in discriminatory behaviour, the perpetration of hate crimes, and so on. Although the speaker bears primary fault for misleading hearers, to the extent that hearers behave epistemically recklessly, failing to question which contextual reference point is being presupposed, their behaviour may also be subject to criticism. As Langton (2018) has argued, there are instances in which accommodating a presupposition can itself be a harmful act.

4.3. Indifference

In the previous section, I considered cases where speakers actively believe reference point information to be false (or to presuppose something false). However, it is also possible that a speaker could intend to convey reference point information while simply lacking *any* beliefs about its truth or falsity. These cases, which are akin to ‘bullshitting’ in the sense of Frankfurt (2005), would violate Grice’s second Quality maxim: the speaker does not have adequate evidence for her assertion.

As before, where hearers can reasonably expect the speaker to be bullshitting, it would be epistemically reckless for them to infer that the speaker believes the reference point information. At the same time, of course, speakers who bullshit are also at fault for violating an important communicative norm. As discussed in the previous section, that norm does not apply to any lesser degree to a speaker who violates it *predictably*. Therefore, the fault attaching to the speaker will not be mitigated by any epistemic recklessness on the part of the hearer.

To conclude this section, I have argued that where hearers have reasonable expectations that the speaker is adhering to the Quality maxims, it is epistemically rational to attribute to the speaker a *belief* that the reference point information obtains (and not merely an intention to communicate that information). Where hearers have reasonable expectations that the speaker is violating Quality norms, it is epistemically irrational to attribute such a belief. Nevertheless,

whenever a speaker violates Quality, her culpability for doing so remains unmitigated by any epistemic recklessness on the hearer's part.

5. Reliability of Beliefs

Having considered whether, and when, hearers may rationally attribute to a speaker a *belief* in reference point information, I turn now to consider whether and when they may rationally infer that information to be *true*.²²⁵ I will argue that this depends, *inter alia*, on whether the speaker's belief can reasonably be expected to be reliable – i.e. whether it is likely to track the truth.²²⁶ Broadly speaking, where hearers have reason to think that the speaker's beliefs will be reliable, it is rational to treat the reference point information as true. Conversely, where hearers have reason to expect the speaker's beliefs to be *unreliable*, that would be irrational. Echoing the discussion in §4, there are a couple of ways this might play out, which are discussed in turn below.

5.1. False Beliefs

In the first kind of case, the reference point information believed by the speaker might turn out to be straightforwardly false. For example, a speaker might utter (2) believing and intending to communicate (4) whereas, in fact, the player missed only a relatively small proportion of his shots.

(2) This player missed 40% of his shots of his shots last season.

(4) This player missed a relatively large proportion of his shots last season.

There could be good reasons for hearers to expect the speaker's belief to be unreliable – say because the speaker is known to be acquiring information from a dubious source. If so, it would clearly be epistemically irrational for hearers to infer the reference point information. At the same time, the speaker may also be at fault, for example because relying on dubious sources contravenes epistemic norms to seek out reliable information. Of course, various factors may affect the extent to which we hold speakers culpable for poor epistemic practices: for example, as high quality sources become easier to access, the failure to make use of them seems increasingly epistemically negligent. It is also possible to imagine cases where the speaker exhibits perfectly good epistemic practices while still ending up with false beliefs, due to factors outside her control. In these cases, I suggest, the speaker attracts no fault (and hearers need not do so either).

²²⁵ Harris (2019) describes this step in successful communication as 'acceptance'.

²²⁶ It will also depend on the prior probability of the reference point information obtaining. However, I will set this consideration aside here.

5.2. False Presuppositions

In other cases of unreliability, the reference point the speaker has in mind may fail to correspond to the one that is salient to the audience.²²⁷ To illustrate, consider an exchange that took place on Twitter on 9th September 2019. The Scottish Liberal Democrats initially tweeted the following:



More than 10% of Scottish pupils are being taught in schools deemed poor or bad quality.

It's disgraceful that the condition of our schools could distract from learning and prove unpleasant workplaces for teachers.



I will focus on the following statement:

(14) More than 10% of Scottish pupils are being taught in schools deemed poor or bad quality.

Plausibly, the author of the tweet believed, and intended to convey, the following reference point information:

(15) A relatively large proportion of Scottish pupils are being taught in schools deemed poor or bad quality.

The reference point which the author had in mind may be the desired situation in which *no* Scottish pupils are being taught in schools deemed poor or bad quality (because all are being taught in good schools). However, the Scottish Nationalist First Minister of Scotland, Nicola Sturgeon, replied to the tweet by pointing out that the proportion stood at almost 40% when the Scottish Liberal Democrats left office in 2007:

²²⁷ Note how this differs from the scenario discussed in §4, where the speaker had *no* reference point in mind.



One gloss of what is going on here is that Sturgeon recognised the salient reference point to be the *prior* proportion of children being taught in schools deemed poor or bad quality. Thus, she took (14) to implicate (16), which is false:

(16) A larger proportion of Scottish pupils are now being taught in schools deemed poor or bad quality than previously (when the Scottish Lib Dems were in office).

Assuming that Sturgeon is right about which reference point is salient, such that audiences of the initial tweet will typically infer (16), the Scottish Lib Dems might be considered to have conveyed unreliable reference point information (whether deliberately or inadvertently).²²⁸ To the extent that they should have known what the salient reference point was they may incur some fault for hearers inferring (16). On the other hand, hearers are surely also well advised to consider the probable political biases inherent in such communications, rather than taking them at face value.

To conclude this section, if the reliability condition is met, the hearer is epistemically rational in coming to believe that the reference point information is *true* (not just that the speaker believes it). If, on the other hand, the speaker's beliefs can be expected to be unreliable, in one or both of the ways described here, the hearer would be epistemically irrational to treat the reference point information as true. Nevertheless, I have suggested that the speaker could incur fault as well, insofar as her unreliability results from poor epistemic practices.

²²⁸ Of course, there are other possible interpretations. For example, both reference points might have been equally salient, or Sturgeon's response could be understood as *changing* the reference point in order to score a political point.

6. Task-Relevance

Although I have argued that a hearer can rationally represent a situation as being different under alternative frames, this doesn't automatically mean the hearer should *respond* differently; that will depend on whether the difference in the represented situation matters to the task at hand. Finally, then, if we are to provide a fully rationalising explanation of framing effects, hearers must reasonably take reference point information to be *relevant*.²²⁹ The proponents of the reference point hypothesis claim that reference points are typically relevant, writing:

Whenever considerations of relative state matter (and they usually do), a population of rational agents, cognizant of the impact of reference points on frame selection, would exhibit a valence-consistent shift. (Sher & McKenzie, 2006, p. 482)

I will show that the situation is a little more nuanced than is suggested here. I begin by teasing out three distinct claims from the quote above:

- i) Reference point information *can* be relevant for making judgements.
- ii) Reference point information is *usually* relevant for making judgements.
- iii) It is rational for hearers to respond *differently* to distinct reference point information.

As discussed in §2, claim (i) seems clearly right – comparative information *can* be relevant to a judgement task. For example, in evaluating the performance of a basketball player, it is necessary to compare that player against others, or against some standard or benchmark.²³⁰ Claim (ii) is harder to assess, since it is an empirical question whether or not reference point information is *usually* relevant. Although I do not believe Sher and McKenzie show that it is, I will set the point aside for now in order to focus on claim (iii).

I want to point out, first, that not all reference points are equally relevant, and therefore it is not always rational for hearers to allow reference point information to affect their judgements to the same degree. For example, if one is choosing players for a team, information about the *best* player in the pool may be less relevant than information about the *average* player. Therefore, if the best player happened to be the contextually salient reference point (say because s/he was

²²⁹ This step corresponds to what Harris dubs 'influence', whereby the hearer enters further mental states, or takes further actions, partly as a result of accepting the information the speaker intended to communicate (Harris, 2019).

²³⁰ At the extremes, some evaluative judgements may be possible without any comparative information: for example, it is obvious that a player who makes 100% of his shots can't count as making a relatively small proportion (even if it is not yet possible to tell whether he counts as making a relatively large proportion). Conversely, a player who misses 100% can't count as making a relatively large proportion. However, for levels of performance in between the extremes, it is difficult to make a sensible judgement without a benchmark.

recently discussed) a speaker's use of the 'missed' frame (2) to describe another player should not necessarily lead the hearer to infer that that player is not good enough for the team.

A real-world example of relevance failure comes from a beer mat issued by Wetherspoons in August 2018:²³¹



Point number 5 here states:

If the unelected President Juncker and his apparatchiks continue to be obstructive, remember that all EU products can be replaced by similar alternatives from the UK – or from the 93% of the world not in the EU.

²³¹ The image is taken from the Full Fact website: <https://fullfact.org/europe/beer-mat-brexit-do-facts-stack/> (accessed on 28th September 2019).

I focus here on the following claim:

(17) 93% of the world is not in the EU.

According to the reference point hypothesis, (17) will typically convey the following information:

(18) A relatively large proportion of the world is outside the EU.

The propositions expressed by (17) and (18) are true with respect to the distribution of the global population. However, this seems largely irrelevant to the point being made about products currently imported from the EU. Whether these products could be imported from outside the EU depends on factors like whether the products are being produced there, whether the producers are willing to export them, whether they can be imported under existing regulations, and so on. In this case, then, hearers would be irrational to place much weight on the reference point information in (18) in judging the UK's trade opportunities post-'Brexit'.

The second point I want to make about Sher and McKenzie's claim (iii) is that it assumes the identity of the reference point is held fixed across framing conditions. Consider again frames (1) and (2):

(1) This player made 40% of his shots last season.

(2) This player missed 60% of his shots last season.

If we assume that the salient reference point under each frame is the *average* player in the domain, the frames will convey the following reference point information, respectively:

(19) This player made a larger proportion of his shots last season than the average player.

(20) This player missed a larger proportion of his shots last season than the average player.

On that basis, it can certainly be rational for hearers to evaluate the player more favourably under frame (1) than frame (2). Things become less clear, though, if each frame invokes a *different* reference point. Consider the following scenario, which is a modified version of the one used by Leong et al. (2017):

Imagine that you are a recruiter for a college basketball team. Your job is to search for a promising high school basketball player and try to recruit him to your college. You are looking through files for players from local high schools, and you are especially interested in players who can score many points. You check in with two recruiters from other colleges in your league. Sam has recruited a player who made 50% and missed

50% of his shots last season. Alex has recruited a player who made 70% and missed 30% of his shots last season.

Hearers would then receive one or other of the following framing conditions:

Framing condition 1.

You are currently looking at a player's file. This player made 60% of his shots last season. How valuable do you think this player would be to your basketball team?

Framing condition 2.

You are currently looking at a player's file. This player missed 40% of his shots last season. How valuable do you think this player would be to your basketball team?

The vignette makes two reference points available – Sam's player, who made 50% and missed 50% of his shots, and Alex's player, who made 70% and missed 30%. Plausibly, then, the two framing conditions could make *different* players salient.²³² After all, the reference point information associated with the 'made' frame presupposes a salient reference point player who made less than 60% of his shots. Therefore, this frame could make Sam's player most salient, giving us the following reference point information:

(21) This player made a larger proportion of his shots last season than the player recruited by Sam.

In contrast, the reference point information associated with the 'missed' frame presupposes a salient reference point player who missed less than 40% of his shots. It may thus serve to make Alex's player most salient, giving us the following reference point information:

(22) This player missed a larger proportion of his shots last season than the player recruited by Alex.

Crucially, (21) and (22) are perfectly consistent with one another. Since they invoke different reference points, they do not rationally require the hearer to evaluate the player differently under each frame (unlike (19) and (20) which invoke the *same* reference point). Therefore, hearers are not automatically rational in responding differently to (1) and (2) in this context. In general terms, they may not be rational to respond differently to alternative frames where those frames bring different reference points to salience.

²³² Of course, this suggestion stands in need of empirical confirmation. For now, though, I simply want to see what would follow if it were correct.

A rationalising explanation is still available, though, even in this scenario. If hearers can reasonably expect the speaker to be invoking the *most relevant* reference point, then it is reasonable to allow the reference point information to inform their judgements. For example, in the scenario described above, let's assume that the 'made' frame conveys (21), which treats Sam's player as the reference point. If the speaker is presumed to be conveying relevant information, then Sam's player must be more relevant than Alex's for some reason (perhaps, say, because Sam is thought to have more realistic standards than Alex). In this framing condition, the target player can be understood to perform *better* than the relevant reference point. Conversely, in the second framing condition, the use of the 'missed' frame is assumed to convey (22), which treats Alex's player as the reference point. If the speaker is presumed to be conveying relevant information, then this time Alex's player must be assumed to be the more relevant reference point. In this case, hearers may infer that the target player performs worse than the relevant reference point, and they are therefore rational to consider him to be a relatively *poor* performer.

To conclude this section, whether the hearer is justified in treating the salient reference point as relevant might depend on whether the speaker can reasonably be presumed to be communicating relevantly, adhering to something like Grice's maxim of Relation. If hearers have good reason to doubt that the speaker is being maximally relevant then it would be epistemically reckless to make use of the reference point information. In other words, it only follows that hearers are rationally relying on reference point information when the *salient* reference point can also be presumed to be the *most relevant*. Once again, though, we should acknowledge that it is not only hearers who can incur fault here. Speakers who (deliberately or inadvertently) communicate irrelevant reference point information may also attract some blame insofar as they are violating the maxim of Relation.

7. Returning to Risky-Choice Framing

I have argued that the information leakage account can provide a rationalising explanation of attribute framing effects, at least in some contexts, albeit the picture is rather more complicated than Sher and McKenzie acknowledge. Before concluding, I want to consider whether the argument generalises to risky-choice framing effects. Let's reconsider the classic 'Asian Disease Problem', reproduced for a final time below:²³³

²³³ I will ignore worries about semantic inequivalence now, which are addressed in earlier chapters.

Imagine the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume the exact scientific estimates of the consequences are as follows:

Framing condition 1:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.

Framing condition 2:

If Program C is adopted, 400 people will die.

If Program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.

According to the reference point hypothesis, the following information will tend to be conveyed by the sure options of the ‘be saved’ and ‘die’ conditions, respectively:²³⁴

(23) If Program A is adopted, a relatively large number of people will be saved.

(24) If Program C is adopted a relatively large number of people will die.

The reference point information is not obviously defeated, so participants are presumably justified in inferring that the experimenters intend to communicate it. Assuming that it is also reasonable for participants to presume that the experimenters are communicating sincerely and reliably, then it may rationally be inferred that the reference point information is true.

One might object here that, since the scenario is clearly hypothetical, the presumption of the speaker’s cooperativity is suspended, meaning that therefore hearers are irrational to infer the reference point information. I don’t think this is quite right, though. First, in order to complete the task sensibly, participants must still trust the experimenters to give an accurate description of the hypothetical scenario, i.e. a description which *would* be true *if* the scenario obtained. Sher and McKenzie make a similar point, writing:

Of course, the signal-selection process in experimental design is very different from those that operate in the “natural world”; but the stimulus-selection process in experiments is typically concealed from the subject’s view, and in many experiments

²³⁴ I take it that reference point information cannot be rationally derived from the *risky* options, since it would be incoherent for *both* probabilities stated there to be large relative to a single contextual reference point.

cooperativeness requires the subject to treat artificial stimuli roughly as they would in corresponding natural situations. (Sher & McKenzie, 2011, p. 43)

Moreover, just because the scenario is hypothetical, it doesn't follow that *everything* the experimenter communicates should be treated as false. Instead, it can be reasonable for participants to treat certain background information as true.²³⁵ In particular, I suggest that the experimenters are reasonably understood as communicating accurate background information about *reference points*.

Turning to issues of relevance, the two framing conditions plausibly make different reference points salient (similarly to the case discussed in §6). The 'be saved' frame intuitively makes salient the expected outcome of all 600 people dying. Meanwhile, the 'missed' frame makes salient the desired outcome of no-one dying. That would explain why the outcome of Program A is understood as a *gain*, whereas the outcome of Program C is understood as a *loss*.²³⁶ Assuming that the experimenters are providing the most relevant information to participants, it is perfectly rational to represent the outcome of Program A as a gain – it genuinely *is* a gain, relative to the relevant reference point. Likewise, it is rational for hearers to represent the outcome of Program C as a loss, since it *is* a loss, relative to the relevant reference point.

Crucially, this would effectively reduce the risky-choice framing effect observed under the ADP to a mere *reflection* effect. As discussed in chapter 1, reflection effects concern *actual* gains and losses, whereas framing effects concern merely *apparent* gains and losses. Of course, the reference point hypothesis has nothing to say about whether *reflection effects* are rational. It might still be irrational to be risk-averse in gain situations and risk-seeking in loss situations; nothing I have said here bears on that question. The crucial point is that the information leakage account can provide a rationalising explanation of that portion of the effect which is attributed to *framing*, i.e. the presentation of prospects using different wording.

8. Conclusion

I believe Sher and McKenzie are right to challenge the standard view of framing effects as evidence of flawed reasoning on the part of 'framed' subjects, given the pragmatic account they put forward. If their information leakage account is correct, hearers do not violate the rational principle of description invariance. Although I have shown that their justification in inferring and using the reference point information will depend on several further conditions obtaining, I

²³⁵ Langton and West (1999) make a similar point with respect to fictional accounts.

²³⁶ Alternatively, it is possible that participants simply accommodate the experimenter's presupposed reference points (as discussed in §4.2), rather than explicitly linking them back to the vignette in the way described here.

have sought to balance their epistemic responsibilities against the communicative responsibilities of speakers.

Summarising the overall thesis, I have argued that a closer focus on the semantics and pragmatics of uttered frames can explain why they are interpreted as describing distinct states of affairs and, therefore, why they provoke distinct responses. This answers the question raised in chapter 1, of *how* and *why* alternative frames induce distinct mental representations. I suggest that this is due to the way our choice of words track genuine differences in the states of the world. As competent language users, we are sensitive to a very intricate network of relations between words and the world.

Recapping the specific claims, I have proposed that framing effects may be attributable to a combination of two factors: lower-bounded interpretations of the quantities denoted by number expressions (as discussed in chapter 4) and the inference of reference point information (as discussed in chapter 6). Although framing effects cannot be traced back to the conventional semantics of the predicate expressions substituted across framing conditions (as argued in chapter 3), their use conveys information that can be analysed as a pragmatic implicature (as discussed in chapters 7). The proposed analysis has important implications for the way in which we think about linguistic effects in philosophy. In chapter 8, I argued that the case study of framing supports a broadly Minimalist approach to semantics and pragmatics (contributing to the philosophical debate discussed in chapters 2 and 5). There are also practical implications. In this final chapter, I have shown how, by applying the theoretical analysis to uses of framing in public discourse, we arrive at a more nuanced understanding of the responsibilities of speakers and hearers.

As I have indicated at various points throughout the thesis, the research I have begun here could fruitfully be extended in various directions. For example, it would be helpful to refine the information leakage account further, to explore its application to risky choice framing, to understand how particular reference points are made salient in context, and to clarify the relationship between reference point information, lower-bounded interpretations of number expressions, and the effects of markedness. Painting with broader brushstrokes, the idea that relative state information bears on our understanding of linguistic stimuli is one which may turn out to have wider philosophical application, well beyond cases of equivalency framing. By bringing the framing literature into closer contact with philosophical debates about the semantics and pragmatics of language, I hope to have paved the way for the further integration of these, along with related discussions happening in the disciplines of philosophy, psychology, linguistics, and further afield.

APPENDIX

Given emerging evidence that the integrity of MTurk-based studies has been compromised by 'bots' or responses from individuals using Virtual Private Servers (VPS) to fake their location, the following screening procedures were performed *ex post* to identify suspicious or low-quality responses in all datasets. These procedures made use of a filtering script available at the following link: <https://osf.io/2uxk9/>.

Using an approach outlined by Dennis, Goodson, and Pearson (forthcoming), MTurk responders' GPS locations were filtered and identical locations were flagged. Any identical GPS locations with unique IP addresses were then further analysed. Each IP address was traced (through iplocation.net) allowing us to identify whether the response came from an ISP or a VPS provider, and whether it came from a server farm. If responses were found to derive from a server farm, an open-ended question in each dataset was then screened by the experimenter to check the integrity and quality of the response using five criteria adapted from (Dennis et al., forthcoming):

- 1) Whether the response is completely unrelated to the study
- 2) Whether the response is incoherent and/or nonsensical
- 3) Whether the response includes a gross misuse of the English language
- 4) Whether the response clearly articulates reasoning behind decisions
- 5) Whether the response cites case facts

Using this procedure, in the first experiment, 46 responses in the dataset were flagged. Of these, 17 had already been screened out for other reasons. Of the remaining 29 responses, 10 were removed on the basis of the answers provided to the open-ended question, or because they were probable duplicates.

In the second experiment, 103 responses in the dataset were flagged. Of these, 45 had already been screened out for other reasons. Of the remaining 58 responses, 30 were removed on the basis of the answers provided to the open-ended question, or because they were probable duplicates.

BIBLIOGRAPHY

- Bach, K. (1994). Conversational Implicature. *Mind & Language*, 9(2), 124-162. doi:10.1111/j.1468-0017.1994.tb00220.x
- Bach, K. (2001). You Don't Say? *Synthese*, 128(1/2), 15-44. doi:10.1023/A:1010353722852
- Bach, K. (2006a). The excluded middle: Semantic minimalism without minimal propositions. *Philosophy and Phenomenological Research*, 73(2), 435-442.
- Bach, K. (2006b). The top 10 misconceptions about implicature. In B. J. Birner & G. Ward (Eds.), *Drawing the Boundaries of Meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn* (pp. 21-30). Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Beaver, D. I., & Geurts, B. (2014). Presupposition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition ed.).
- Berkeley, D., & Humphreys, P. (1982). Structuring Decision Problems and the 'Bias Heuristic'. *Acta Psychologica*, 50, 201-252.
- Bermúdez, J. L. (2009). The Second Challenge: Individuating Outcomes. In J. L. Bermúdez (Ed.), *Decision Theory and Rationality* (pp. 77-107). Oxford: Oxford University Press.
- Bermúdez, J. L. (2018). Frames, Rationality, and Self-Control. In J. L. Bermúdez (Ed.), *Self-Control, Decision Theory, and Rationality* (pp. 179-203). Cambridge: Cambridge University Press.
- Bezuidenhout, A. L. (1997). The Communication of "De Re" Thoughts. *Noûs*, 31(2), 197.
- Bohm, P., & Lind, H. (1992). A note on the robustness of a classical framing result. *Journal of Economic Psychology*, 13(2), 355-361. doi:10.1016/0167-4870(92)90039-A
- Bolinger, R. J. (2017). The Pragmatics of Slurs. *Noûs*, 51(3), 439-462. doi:10.1111/nous.12090
- Borg, E. (2004). *Minimal Semantics*. Oxford: Clarendon Press.
- Borg, E. (2007). Minimalism versus Contextualism in Semantics. In G. Preyer & G. Peter (Eds.), *Context-Sensitivity and Semantic Minimalism: New Essays on Semantics and Pragmatics*. Oxford: Oxford University Press.
- Borg, E. (2009). On Three Theories of Implicature: Default Theory, Relevance Theory and Minimalism. *International Review of Pragmatics*, 1(1), 63-83. doi:10.1163/187731009X455848
- Borg, E. (2012). *Pursuing Meaning*. Oxford: Oxford University Press.
- Borg, E. (2016). Exploding Explicatures. *Mind & Language*, 31(3), 335-355. doi:10.1111/mila.12109
- Borg, E. (2017). Explanatory Roles for Minimal Content. *Noûs*. doi:10.1111/nous.12217
- Borg, E., & Connolly, P. (forthcoming). Exploring Linguistic Liability. In E. Lepore & D. Sosa (Eds.), *Oxford Studies in Philosophy of Language*. Oxford: Oxford University Press.
- Borg, E., & Fisher, S. A. (forthcoming). Semantic Content and Utterance Context: A Spectrum of Approaches. In P. Stalmaszczyk (Ed.), *Philosophy of Language: Cambridge Handbooks in Language and Linguistics*. Cambridge: Cambridge University Press.
- Brainerd, C. J., & Kingma, J. (1984). Do Children Have to Remember to Reason? A Fuzzy-Trace Theory of Transitivity Development. *Developmental Review*, 4, 311-377.
- Brainerd, C. J., & Kingma, J. (1985). On the Independence of Short-Term Memory and Working Memory in Cognitive Development. *Cognitive Psychology*, 17, 210-247.
- Breheny, R. (2008). A New Look at the Semantics and Pragmatics of Numerically Quantified Noun Phrases. *Journal of Semantics*, 25(2), 93-139.
- Cappelen, H., & Lepore, E. (2002). Indexicality, Binding, Anaphora and A Priori Truth. *Analysis*, 62(4), 271-281. doi:10.1093/analysis/62.4.271
- Cappelen, H., & Lepore, E. (2005). *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Oxford: Wiley-Blackwell.
- Cappelen, H., & Lepore, E. (2006). Replies. *Philosophy and Phenomenological Research*, 73(2), 469-492. doi:10.1111/j.1933-1592.2006.tb00630.x
- Carston, R. (1988). Implicature, Explicature, and Truth-Theoretic Semantics. In R. M. Kempson (Ed.), *Mental representations: The interface between language and reality* (pp. 155-181): Cambridge University Press.
- Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance Theory: Applications and implications*: John Benjamins Publishing Company.

- Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.
- Carston, R. (2009). Relevance Theory: Contextualism or Pragmaticism? *UCL Working Papers in Linguistics*, 21, 17-24.
- Carston, R. (2012). Word meaning and concept expressed. *The Linguistic Review*, 29(4), 607. doi:10.1515/tlr-2012-0022
- Carston, R. (2013). Word Meaning, What is Said, and Explicature. In C. Penco & F. Domaneschi (Eds.), *What is Said and What is Not*: Stanford: Csl Publications.
- Carston, R. (2019). Ad Hoc Concepts, Polysemy and the Lexicon. In B. Clark, K. Scott, & R. Carston (Eds.), *Relevance, Pragmatics and Interpretation* (pp. 150-162). Cambridge: Cambridge University Press.
- Carston, R., & Hall, A. (2012). Implicature and Explicature. In H.-J. Schmid (Ed.), *Cognitive Pragmatics* (pp. 47-84).
- Chick, C. F., Reyna, V. F., & Corbin, J. C. (2016). Framing effects are robust to linguistic disambiguation: A critical test of contemporary theory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 42(2), 238-256. doi:10.1037/xlm0000158
- Chierchia, G., Fox, D., & Spector, B. (2012). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In C. Maienborn, K. von Stechow, & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning* (Vol. 3, pp. 2297-2332). Berlin: Mouton de Gruyter.
- Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Collins, J. (2007). Syntax, More or Less. *Mind*, 116(464), 805-850.
- Corner, A., & Hahn, U. (2010). Message Framing, Normative Advocacy and Persuasive Success. *Argumentation*, 24(2), 153-163. doi:10.1007/s10503-009-9159-6
- Davis, W. A. (1998). *Implicature: Intention, Convention, and Principle in the Failure of Gricean Theory*. Cambridge: Cambridge University Press.
- Davis, W. A. (2016). Implicature. In *Oxford Handbooks Online*. doi:10.1093/oxfordhb/9780199935314.013.21
- Degen, J., & Tanenhaus, M. K. (2015). Processing Scalar Implicature: A Constraint-Based Approach. *Cognitive Science*, 39(4), 667-710. doi:10.1111/cogs.12171
- Dennis, S. A., Goodson, B. M., & Pearson, C. (forthcoming). Online Worker Fraud and Evolving Threats to the Integrity of MTurk Data: A Discussion of Virtual Private Servers and the Limitations of IP-Based Screening Procedures. *Behavioral Research in Accounting*. doi:10.2139/ssrn.3233954
- Druckman, J. N. (2004). Political preference formation: competition, deliberation, and the (Ir)relevance of Framing Effects. *American Political Science Review*, 98(4), 671-686.
- Druckman, J. N. (2011). What's It All About?: Framing in Political Science. In G. Keren (Ed.), *Perspectives on Framing* (pp. 279-301). Abingdon, Oxfordshire; New York: Psychology Press.
- Dummett, M. A. E. (1978). *Truth and Other Enigmas*. Cambridge, MA: Harvard University Press.
- Dummett, M. A. E. (1981). *Frege: Philosophy of Language* (2 ed.). London: Duckworth.
- Evans, G. (1982). *The Varieties of Reference* (J. McDowell Ed.). Oxford: Clarendon Press.
- Fagley, N. S. (1993). A Note Concerning Reflection Effects Versus Framing Effects. *Psychological Bulletin*, 113(3), 451-452. doi:10.1037/0033-2909.113.3.451
- Fisher, S. A. (2019). Reassessing truth-evaluability in the Minimalism-Contextualism debate. *Synthese*, 1-18. doi:10.1007/s11229-019-02245-2
- Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford; New York: Clarendon Press.
- Frankfurt, H. G. (2005). *On Bullshit*. Princeton, NJ: Princeton University Press.
- Frege, G. (1948). Sense and Reference. *The Philosophical Review*, 57, 209-230.
- Frisch, D. (1993). Reasons for Framing Effects. *Organizational Behavior and Human Decision Processes*, 54(3), 399-429. doi:10.1006/obhd.1993.1017

- García-Carpintero, M. (1998). Indexicals as Token-Reflexives. *Mind*, 107(427), 529-563. doi:10.1093/mind/107.427.529
- García-Carpintero, M. (2013). Pursuing Meaning, by Emma Borg. *Mind*, 122(486), 515-521. doi:10.1093/mind/fzt076
- Gauker, C. (1998). What Is a Context of Utterance? *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 91(2), 149-172. doi:10.1023/A:1004247202476
- Gauker, C. (2008). Zero Tolerance for Pragmatics. *Synthese*, 165(3), 359-371. doi:10.1007/s11229-007-9189-2
- Gauker, C. (2019). Against the speaker-intention theory of demonstratives. *Linguistics and Philosophy*, 42, 109-129.
- Geurts, B. (1998). Presuppositions and Anaphors in Attitude Contexts. *Linguistics and Philosophy*, 21(6), 545-601. doi:10.1023/A:1005481821597
- Geurts, B. (2006). Take "five": the meaning and use of a number word. In S. Vogeeler & L. Tasmowski (Eds.), *Non-definiteness and plurality* (pp. 311-329). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Geurts, B. (2010). *Quantity Implicatures*. Cambridge: Cambridge University Press.
- Geurts, B. (2013). Alternatives in framing and decision making. *Mind and Language*, 28(1), 1-19. doi:10.1111/mila.12005
- Geurts, B., & Rubio-Fernández, P. (2015). Pragmatics and Processing. *Ratio*, 28, 446-469.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, 62(1), 451-482. doi:10.1146/annurev-psych-120709-145346
- Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA; London: Harvard University Press.
- Hall, A. (2008). Free Enrichment or Hidden Indexicals? *Mind and Language*, 23(4), 426-456.
- Hansen, N. (2014). Contrasting cases. In J. Beebe (Ed.), *Advances in experimental epistemology*. London: Bloomsbury.
- Hansen, N., & Chemla, E. (2013). Experimenting on Contextualism. *Mind & Language*, 28(3), 286-321. doi:10.1111/mila.12019
- Harris, D. (2019). We talk to people, not contexts. *Philosophical Studies*. doi:10.1007/s11098-019-01335-8
- Harris, D. (forthcoming). Semantics without Semantic Content. *Mind & Language*.
- Higginbotham, J., & Segal, G. (1994). Priorities in the Philosophy of Thought. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 68(1), 85-130. doi:10.1093/aristoteliansupp/68.1.85
- Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin*, 118(2), 248-271. doi:10.1037//0033-2909.118.2.248
- Holleman, B. C., & Pander Maat, H. L. W. (2009). The pragmatics of profiling: Framing effects in text interpretation and text production. *Journal of Pragmatics*, 41(11), 2204-2221. doi:10.1016/j.pragma.2008.09.038
- Honda, H., & Yamagishi, K. (2017). Communicative functions of directional verbal probabilities: Speaker's choice, listener's inference, and reference points. *Quarterly Journal of Experimental Psychology*, 70(10), 2141-2158. doi:10.1080/17470218.2016.1225779
- Horn, L. (1989). *A Natural History of Negation*. Chicago: University of Chicago Press.
- Hsee, C. K., Blount, S., Loewenstein, G. F., & Bazerman, M. H. (1999). Preference Reversals Between Joint and Separate Evaluations of Options: A Review and Theoretical Analysis. *Psychological Bulletin*, 125(5), 576-590. doi:10.1037/0033-2909.125.5.576
- Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, 58(3), 376-415. doi:10.1016/j.cogpsych.2008.09.001
- Ingram, J., Hand, C. J., & Moxey, L. M. (2014). Processing inferences drawn from the logically equivalent frames half full and half empty. *Journal of Cognitive Psychology*, 26(7), 799-817. doi:10.1080/20445911.2014.956747
- Ingram, J., & Moxey, L. M. (2011). Complement set focus without explicit quantity. *Journal of Cognitive Psychology*, 23(3), 383-400. doi:10.1080/20445911.2011.524202
- Jou, J., Shanteau, J., & Harris, R. J. (1996). An information processing view of framing effects: The role of causal schemas in decision making. *Memory and Cognition*, 24(1), 1-15. doi:10.3758/BF03197268

- Kahneman, D. (2000). Preface. In D. Kahneman & A. Tversky (Eds.), *Choices, values, and frames*. Cambridge: Cambridge University Press.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. London: Penguin.
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-292.
- Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, 11(2), 123-141. doi:[https://doi.org/10.1016/0010-0277\(82\)90022-1](https://doi.org/10.1016/0010-0277(82)90022-1)
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39(4), 341-350. doi:10.1037//0003-066X.39.4.341
- Kaplan, D. (1989). Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and other Indexicals. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes From Kaplan* (pp. 481-563). Oxford: Oxford University Press.
- Karevold, K. I., & Teigen, K. H. (2010). Progress framing and sunk costs: How managers' statements about project progress reveal their investment intentions. *Journal of Economic Psychology*, 31, 719-731.
- Kempson, R. M., & Cormack, A. (1981). Ambiguity and Quantification. *Linguistics and Philosophy*, 4(2), 259-309.
- Kennedy, C. (2007). Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives. *Linguistics and Philosophy*, 30(1), 1-45. doi:10.1007/s10988-006-9008-0
- Kennedy, C., & McNally, L. (2005). Scale Structure, Degree Modification, and the Semantics of Gradable Predicates. *Language*, 81(2), 345-381. doi:10.1353/lan.2005.0071
- Keren, G. (2011). On the Definition and Possible Underpinnings of Framing Effects: A Brief Review and a Critical Evaluation. In G. Keren (Ed.), *Perspectives on Framing* (pp. 3-33). Abingdon, Oxfordshire; New York: Psychology Press.
- King, J. C., & Stanley, J. (2005). Semantics, pragmatics, and the role of semantic content. In Z. G. Szabo (Ed.), *Semantics Versus Pragmatics* (pp. 111--164). Oxford: Oxford University Press.
- Kratzer, A. (2012). *Modals and Conditionals: New and Revised Perspectives*. Oxford: Oxford University Press.
- Krifka, M. (2007). Negated antonyms: Creating and filling the gap. In U. Sauerland & P. Stateva (Eds.), *Presupposition and Implicature in Compositional Semantics* (pp. 163--177); Palgrave-Macmillan.
- Kühberger, A. (1995). The Framing of Decisions: A New Look at Old Problems. *Organizational Behavior and Human Decision Processes*, 62(2), 230-240. doi:<https://doi.org/10.1006/obhd.1995.1046>
- Kühberger, A. (1998). The Influence of Framing on Risky Decisions: A Meta-analysis. *Organizational Behavior and Human Decision Processes*, 75(1), 23-55. doi:10.1006/obhd.1998.2781
- Kühberger, A. (2002). The Rationality of Risky Decisions: A Changing Message. *Theory & Psychology*, 12(4), 427-452. doi:10.1177/0959354302012004293
- Kühberger, A., & Gradl, P. (2013). Choice, Rating, and Ranking: Framing Effects with Different Response Modes. *Journal of Behavioral Decision Making*, 26(2), 109-117. doi:10.1002/bdm.764
- Kühberger, A., Schulte-Mecklenbeck, M., & Perner, J. (1999). The Effects of Framing, Reflection, Probability, and Payoff on Risk Preference in Choice Tasks. *Organizational Behavior and Human Decision Processes*, 78(3), 204-231. doi:10.1006/obhd.1999.2830
- Kühberger, A., & Tanner, C. (2010). Risky choice framing: Task versions and a comparison of prospect theory and fuzzy-trace theory. *Journal of Behavioral Decision Making*, 23(3), 314-329. doi:10.1002/bdm.656
- Langton, R. (2018). The Authority of Hate Speech. In J. Gardner, L. Green, & B. Leiter (Eds.), *Oxford Studies in Philosophy of Law* (Vol. 3, pp. 123-152). Oxford: Oxford University Press.
- Langton, R., & West, C. (1999). Scorekeeping in a pornographic language game. *Australasian Journal of Philosophy*, 77(3), 303-319. doi:10.1080/00048409912349061
- Lasersohn, P. (2005). Context Dependence, Disagreement, and Predicates of Personal Taste. *Linguistics and Philosophy*, 28(6), 643-686. doi:10.1007/s10988-005-0596-x
- Lehrer, A. (1985). Markedness and antonymy. *Journal of Linguistics*, 21(2), 397-429. doi:10.1017/S002222670001032X

- Leong, L. M., McKenzie, C. R. M., Sher, S., & Müller-Trede, J. (2017). The Role of Inference in Attribute Framing Effects. *Journal of Behavioral Decision Making*, 30(5), 1147-1156. doi:10.1002/bdm.2030
- Lepore, E., & Stone, M. (2015). *Imagination and Convention: Distinguishing Grammar and Inference in Language*. Oxford: Oxford University Press.
- Levin, I. P. (1987). Associative effects of information framing. *Bulletin of the Psychonomic Society*, 25(2), 85-86.
- Levin, I. P., & Gaeth, G. J. (1988). How Consumers Are Affected By The Framing Of Attribute Info. *Journal of Consumer Research*, 15(3), 374.
- Levin, I. P., Johnson, R. D., Deldin, P. J., Carstens, L. M., Cressey, L. J., & Davis, C. R. (1986). Framing Effects in Decisions with Completely and Incompletely Described Alternatives. *Organizational Behavior and Human Decision Processes*, 38, 48-64.
- Levin, I. P., Johnson, R. D., Russo, C. P., & Deldin, P. J. (1985). Framing Effects in Judgment Tasks with Varying Amounts of Information. *Organizational Behavior and Human Decision Processes*, 36, 362-377.
- Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). Not all frames are created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, 76(2), 149.
- Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA; London: MIT Press.
- Lewis, D. (1979). Scorekeeping in a Language Game. *Journal of Philosophical Logic*, 8(1), 339-359. doi:10.1007/BF00258436
- Macdonald, R. R. (1986). Credible conceptions and implausible probabilities. *The British journal of mathematical and statistical psychology*, 39(1), 15-27. doi:10.1111/j.2044-8317.1986.tb00842.x
- MacFarlane, J. (2007). Semantic Minimalism and Nonindexical Contextualism. In G. Preyer & G. Peter (Eds.), *Context-Sensitivity and Semantic Minimalism: New Essays on Semantics and Pragmatics* (pp. 240-250). Oxford: Oxford University Press.
- MacFarlane, J. (2009). Nonindexical Contextualism. *Synthese*, 166(2), 231-250. doi:10.1007/s11229-007-9286-2
- MacFarlane, J. (2014). *Assessment Sensitivity: Relative Truth and its Applications*. Oxford: Oxford University Press.
- Mandel, D. R. (2001). Gain-Loss Framing and Choice: Separating Outcome Formulations from Descriptor Formulations. *Organizational Behavior and Human Decision Processes*, 85(1), 56-76. doi:10.1006/obhd.2000.2932
- Mandel, D. R. (2013). At least I replied: Reply to Simmons & Nelson's Data Colada blog post. Retrieved from <https://sites.google.com/site/themandelian/data-colada>
- Mandel, D. R. (2014). Do framing effects reveal irrational choice? *Journal of experimental psychology. General*, 143(3), 1185-1198. doi:10.1037/a0034207
- Maule, A. J. (1989). Positive and negative decision frames: A verbal protocol analysis of the Asian Disease problem of Tversky and Kahneman. In H. Montgomery & O. Svenson (Eds.), *Process and structure in human decision making*. Chichester, UK: Wiley.
- McKenzie, C. R. M. (2004). Framing effects in inference tasks—and why they are normatively defensible. *Memory & Cognition*, 32(6), 874-885. doi:10.3758/BF03196866
- McKenzie, C. R. M., & Nelson, J. D. (2003). What a Speaker's Choice of Frame Reveals: Reference Points, Frame Selection, and Framing Effects. *Psychonomic Bulletin & Review: A Journal of the Psychonomic Society, Inc*, 10(3), 596.
- McKenzie, C. R. M., & Sher, S. (2008). Framing effects and rationality. In N. Chater & M. Oaksford (Eds.), *The Probabilistic Mind: Prospects for Bayesian Cognitive Science* (pp. 79-96). Oxford: Oxford University Press.
- Meyerowitz, B. E., & Chaiken, S. (1987). The Effect of Message Framing on Breast Self-Examination Attitudes, Intentions, and Behavior. *Journal of Personality and Social Psychology*, 52(3), 500-510. doi:10.1037/0022-3514.52.3.500
- Moxey, L. (2011). Mechanisms Underlying Linguistic Framing Effects. In G. Keren (Ed.), *Perspectives on Framing* (pp. 119-134). Abingdon, Oxfordshire; New York: Psychology Press.
- Moxey, L., & Sanford, A. J. (1987). Quantifiers and focus. *Journal of Semantics*, 5, 189-206.

- Neale, S. (1992). Paul Grice and the Philosophy of Language. *Linguistics and Philosophy*, 15(5), 509-559. doi:10.1007/BF00630629
- Neale, S. (2007a). Heavy Hands, Magic, and Scene-Reading Traps. *European Journal of Analytic Philosophy*, 3(2), 77-132.
- Neale, S. (2007b). On Location. In M. O'Rourke & C. Washington (Eds.), *Essays on the Philosophy of John Perry* (pp. 251-393). Cambridge MA: MIT Press.
- Neale, S. (2016). Silent Reference. In G. Ostertag (Ed.), *Meanings and Other Things: Essays in Honor of Stephen Schiffer* (pp. 229-344). Oxford: Oxford University Press.
- Okder, H. (2012). The illusion of the framing effect in risky decision making. *Journal of Behavioral Decision Making*, 25(1), 63-73. doi:10.1002/bdm.715
- Panizza, D., Huang, Y. T., Chierchia, G., & Snedeker, J. (2015). Relevance of Polarity for the Online Interpretation of Scalar Terms. *Semantics and Linguistic Theory*, 360. doi:10.3765/salt.v0i0.2530
- Partee, B. (1987). Noun-phrase interpretation and type-shifting principles. In J. Groenendijk, D. de Jong, & M. Stokhof (Eds.), *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers* (pp. 115-144). Dordrecht: Reidel.
- Perry, J. (1986). Thought without Representation. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 60, 137-151.
- Perry, J. (2012). *Reference and Reflexivity* (2 ed.). Stanford, CA: CSLI Publications.
- Pietroski, P. M. (2005). Meaning before Truth. In G. Preyer & G. Peter (Eds.), *Contextualism in Philosophy: Knowledge, Meaning, and Truth*. Oxford: Clarendon Press.
- Pietroski, P. M. (2015). Framing Event Variables. *Erkenntnis*, 80(S1), 31-60. doi:10.1007/s10670-014-9659-0
- Pietroski, P. M. (2018). *Conjoining Meanings: Semantics Without Truth Values*. Oxford: Oxford University Press.
- Potts, C. (2015). Presupposition and Implicature. In S. Lappin & C. Fox (Eds.), *The Handbook of Contemporary Semantic Theory* (pp. 168-202): John Wiley & Sons, Ltd.
- Predelli, S. (1998). I am not here now. *Analysis*, 58(258), 107-115.
- Predelli, S. (2005). *Contexts: Meaning, Truth, and the Use of Language*. Oxford: Clarendon Press.
- Predelli, S. (2013). *Meaning without Truth*. Oxford: Oxford University Press.
- Quine, W. V. (1951). Two Dogmas of Empiricism. *Philosophical Review*, 60, 20-43.
- Rayo, A. (2013). A Plea for Semantic Localism. *Noûs*, 47(4), 647-679. doi:10.1111/j.1468-0068.2011.00846.x
- Recanati, F. (1989). The Pragmatics of What is Said. *Mind and Language*, 4(4), 295-329.
- Recanati, F. (2004). *Literal Meaning*. Cambridge: Cambridge University Press.
- Recanati, F. (2010). *Truth-Conditional Pragmatics*. Oxford: Oxford University Press.
- Recanati, F. (2017). Contextualism and Polysemy. *Dialectica*, 71(3), 379-397. doi:10.1111/1746-8361.12179
- Reyna, V. F., & Brainerd, C. J. (1991). Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion. *Journal of Behavioral Decision Making*, 4(4), 249-262. doi:10.1002/bdm.3960040403
- Reyna, V. F., & Brainerd, C. J. (2011). Dual processes in decision making and developmental neuroscience: A fuzzy-trace model. *Developmental Review*, 31(2), 180-206. doi:10.1016/j.dr.2011.07.004
- Reyna, V. F., Chick, C. F., Corbin, J. C., & Hsia, A. N. (2014). Developmental Reversals in Risky Decision Making: Intelligence Agents Show Larger Decision Biases Than College Students. *Psychological Science*, 25(1), 76-84. doi:10.1177/0956797613497022
- Rothschild, D., & Segal, G. (2009). Indexical Predicates. *Mind & Language*, 24(4), 467-493. doi:10.1111/j.1468-0017.2009.01371.x
- Russell, B. (1957). Mr Strawson on Referring. *Mind*, 66(263), 385-389.
- Ruytenbeek, N., Verheyen, S., & Spector, B. (2017). Asymmetric Inference towards the Antonym: Experiments into the Polarity and Morphology of Negated Adjectives. *Glossa: A Journal of General Linguistics*, 2(1), 1-27. doi:10.5334/gjgl.151
- Sadock, J. M. (1978). On Testing for Conversational Implicature. In P. Cole (Ed.), *Syntax and Semantics: Pragmatics* (pp. 281-297). Cambridge, MA: Academic Press.
- Sanford, A. J., Dawydiak, E. J., & Moxey, L. (2007). A Unified Account of Quantifier Perspective Effects in Discourse. *Discourse Processes: A Multidisciplinary Journal*, 44(1), 1-32. doi:10.1080/01638530701285556

- Sanford, A. J., Fay, N., Stewart, A., & Moxey, L. (2002). Perspective in Statements of Quantity, with Implications for Consumer Psychology. *Psychological Science*, *13*(2), 130-134. doi:10.1111/1467-9280.00424
- Saul, J. M. (2001). Wayne Davis, "Implicature: Intention, Convention, and Principle in the Failure of Gricean Theory" (Book Review). *Noûs*, *35*(4), 630-641.
- Saul, J. M. (2002a). Speaker Meaning, What Is Said, and What Is Implicated. *Noûs*, *36*(2), 228-248. doi:10.1111/1468-0068.00369
- Saul, J. M. (2002b). What Is Said and Psychological Reality; Grice's Project and Relevance Theorists' Criticisms. *Linguistics and Philosophy*, *25*(3), 347-372. doi:10.1023/A:1015221313887
- Saul, J. M. (2013). *Lying, Misleading, and What is Said: An Exploration in Philosophy of Language and in Ethics*. Oxford: Oxford University Press.
- Schwarz, N. (1996). *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*: Lawrence Erlbaum Associates, Inc.
- Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition*, *101*(3), 467-494. doi:10.1016/j.cognition.2005.11.001
- Sher, S., & McKenzie, C. R. M. (2011). Levels of Information: A Framing Hierarchy. In G. Keren (Ed.), *Perspectives on Framing* (pp. 35-63). Abingdon, Oxfordshire; New York: Psychology Press.
- Simmons, J., & Nelson, L. (2013). "Exactly": The Most Famous Framing Effect is Robust to Precise Wording. Retrieved from <http://datacolada.org/11>
- Simons, M. (2006). Foundational Issues in Presupposition. *Philosophy Compass*, *1*(4), 357-372. doi:10.1111/j.1747-9991.2006.00028.x
- Spector, B. (2013). Bare Numerals and Scalar Implicatures. *Language and Linguistics Compass*, *7*, 273-294. doi:10.1111/lnc3.12018
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition* (2 ed.). Oxford: Blackwell.
- Stalnaker, R. C. (1999). *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press.
- Stanley, J. (2000). Context and Logical Form. *Linguistics and Philosophy*, *23*(4), 391-434. doi:10.1023/A:1005599312747
- Stanley, J. (2002). Making it Articulated. *Mind & Language*, *17*(1-2), 149-168. doi:10.1111/1468-0017.00193
- Stanley, J. (2005). Semantics in context. In G. Preyer & G. Peter (Eds.), *Contextualism in Philosophy: Knowledge, Meaning, and Truth* (pp. 221--254). Oxford: Oxford University Press.
- Stanley, J., & Szabo, Z. G. (2000). On Quantifier Domain Restriction. *Mind and Language*, *15*(2), 219-261.
- Stanovich, K. E., & West, R. F. (1998). Individual Differences in Framing and Conjunction Effects. *Thinking & Reasoning*, *4*(4), 289-317. doi:10.1080/135467898394094
- Steiger, A., & Kühberger, A. (2018). A Meta-Analytic Re-Appraisal of the Framing Effect. *Zeitschrift für Psychologie*, *226*(1), 45-55. doi:10.1027/2151-2604/a000321
- Stojnić, U., Stone, M., & Lepore, E. (2017). Discourse and logical form: pronouns, attention and coherence. *Linguistics and Philosophy*, *40*(5), 519-547. doi:10.1007/s10988-017-9207-x
- Strawson, P. F. (1950). On Referring. *Mind*, *59*(235), 320-344.
- Szabo, Z. G. (2001). Adjectives in Context. In I. Kenesei & R. M. Harnish (Eds.), *Perspectives on Semantics, Pragmatics, and Discourse: A Festschrift for Ferenc Kiefer* (pp. 119-146): John Benjamins Publishing Company.
- Taschek, W. W. (1992). Frege's Puzzle, Sense, and Information Content. *Mind*, *101*(404), 767-791. doi:10.1093/mind/101.404.767
- Teigen, K. H. (2016). Framing of Numerical Quantities. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell Handbook of Judgement and Decision Making* (pp. 568-589). Chichester, UK: John Wiley & Sons, Ltd.
- Teigen, K. H., & Brun, W. (1999). The Directionality of Verbal Probability Expressions: Effects on Decisions, Predictions, and Probabilistic Reasoning. *Organizational Behavior and Human Decision Processes*, *80*(2), 155-190. doi:10.1006/obhd.1999.2857
- Teigen, K. H., & Karevold, K. I. (2005). Looking back versus looking ahead: framing of time and work at different stages of a project. *Journal of Behavioral Decision Making*, *18*(4), 229-246. doi:10.1002/bdm.502

- Teigen, K. H., & Nikolaisen, M. I. (2009). Incorrect estimates and false reports: How framing modifies truth. *Thinking & Reasoning*, 15(3), 268-293. doi:10.1080/13546780903020999
- Thierman, J. S. (2014). *Double Attribute Frames: Implications for Theory and Practice*. (Doctor of Philosophy). University of California, San Diego,
- Tombu, M., & Mandel, D. R. (2015). When Does Framing Influence Preferences, Risk Perceptions, and Risk Attitudes? The Explicated Valence Account. *Journal of Behavioral Decision Making*, 28(5), 464-476. doi:10.1002/bdm.1863
- Travis, C. (2000). *Unshadowed Thought: Representation in Thought and Language*. Cambridge, MA: Harvard University Press.
- Travis, C. (2006). *Thought's Footing: A Theme in Wittgenstein's Philosophical Investigations*. Oxford: Oxford University Press.
- Travis, C. (2008). *Occasion-Sensitivity: Selected Essays*. Oxford: Oxford University Press.
- Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, 211(4481), 453-458.
- Tversky, A., & Kahneman, D. (1986). Rational Choices and the Framing of Decisions. *The Journal of Business*, 59, S251-S278.
- Van Schie, E. C. M., & Van Der Pligt, J. (1995). Influencing Risk Preference in Decision Making: The Effects of Framing and Salience. *Organizational Behavior and Human Decision Processes*, 63(3), 264-275. doi:10.1006/obhd.1995.1078
- Verheyen, S., Dewil, S., & Égré, P. (2018). Subjectivity in gradable adjectives: The case of tall and heavy. *Mind & Language*, 33(5), 460-479. doi:10.1111/mila.12184
- Von Fintel, K. (2004). Would you believe it? The king of France is back! Presuppositions and truth-value intuitions. In A. Bezuidenhout & M. Reimer (Eds.), *Descriptions and Beyond* (Vol. 269, pp. 315-341). Oxford: Oxford University Press.
- Wang, X. T. (1996). Framing Effects: Dynamics and Task Domains. *Organizational Behavior and Human Decision Processes*, 68, 145-157.
- Wierzbicka, A. (1985). Different cultures, different languages, different speech acts: Polish vs. English. *Journal of Pragmatics*, 9(2), 145-178. doi:10.1016/0378-2166(85)90023-2
- Wilson, D. K., Kaplan, R. M., & Schneiderman, L. J. (1987). Framing of decisions and selections of alternatives in health care. *Social Behaviour*, 2(1), 51-59.