

Evaluative polarity words in risky choice framing

Annika Wallin ^{a,*}, Carita Paradis ^b, Konstantinos V. Katsikopoulos ^{c,d}

^a Lund University Cognitive Science, Lund University, Box 192, 221 00 Lund, Sweden

^b Centre for Languages and Literature, Lund University, Box 201, 221 00 Lund, Sweden

^c Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

^d Naval Postgraduate School, Department of Operations Research, 1411 Cunningham Road, Monterey, CA 93943, USA

Received 1 June 2016; received in revised form 6 September 2016; accepted 6 September 2016

Available online 1 November 2016



Abstract

This article is concerned with how we make decisions based on how problems are presented to us and the effect that the framing of the problem might have on our choices. Current philosophical and psychological accounts of the framing effect in experiments such as the Asian Disease Problem (ADP) concern reference points and domains (gains and losses). We question the importance of reference points and domains. Instead, we adopt a linguistic perspective focussing on the role of the evaluative polarity evoked by the words – negative and positive – used to describe the options in the decision problem. We show that the evaluative polarity of the different wordings in the ADP better explain participants' behaviour than reference points and domains. We propose two models in which the values given to evaluative polarity words (their valence) directly influence the strength of framing. The results indicate that linguistic considerations regarding evaluative polarity have to be considered in relation to the ADP. The account resembles Fuzzy-Trace-Theory but allows for the strength of evaluative polarity to directly affect behaviour. In the discussion, we also assess how evaluative polarity relates to negation, antonyms and the communicative frame within which the choices are presented.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Asian Disease Problem Evaluative polarity; Framing; Risky choice framing; Valence

1. Introduction

In spite of the fact that decision making plays an important role in most professional and everyday situations in which human beings are involved, and in spite of the fact that information about options more often than not are presented through language, decision making is not a topic that has generated much interest in the linguistics literature. In contrast, human decision making has been the topic of a great deal of active research in philosophy, psychology and cognitive science, and research carried out in those areas has primarily been concerned with human performance in decision making in terms of people's needs, preferences and values.

A topic that has received extensive attention in that literature is *framing*, i.e. the way in which information about options is presented to decision makers. This has been shown to have a great impact on the decisions being made. A framing effect is said to occur when people's inferences or preferences change with different, but logically equivalent descriptions of the same decision problem. Despite the fact that the presentation of the options are provided through the medium of

* Corresponding author. Fax: +46 46 2223228.

E-mail addresses: annika.wallin@lucs.lu.se (A. Wallin), carita.paradis@englund.lu.se (C. Paradis), katsikop@mpib-berlin.mpg.de (K.V. Katsikopoulos).

language, the actual effect of the expressions that are used in the communicative act are largely ignored in framing research. This is, however, what this article is about. More precisely, we investigate the effect the actual words used to present options have on decisions made in [Tversky's and Kahneman's \(1981\) Asian Disease Problem](#) (henceforth, ADP). The ADP provides participants with a cover story about a disease, which is expected to kill a given number of people. In the original ADP experiment, two groups of participants were asked to choose between two programs that have equal expected value for combating the disease as follows.

Cover story

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people.

Version 1

- If Program A is adopted, 200 people will be saved.
- If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.

Version 2

- If Program A is adopted, 400 people will die.
- If Program B is adopted, there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die.

One of the groups of participants was asked to make a decision on the basis of Version 1 and the other group on the basis of Version 2. [Tversky and Kahneman \(1981\)](#) found that 72% of participants chose Program A in Version 1, while 78% of participants chose Program B in Version 2. This result is surprising in the sense that the numerical outcomes of the variants are identical in both versions; they only differ with respect to how the information is presented, i.e. *framing*. In the literature, the framing effect is often explained with reference to Prospect theory, which is an empirically based formal theory. Prospect theory states that people are risk seeking when they encounter potential losses and risk averse when they encounter possible gains. This raises the question of why Version 1 and Version 2 differ with respect to whether they are potential losses or potential gains since the outcomes are identical. Rather than appealing to the domains of gains and losses and reference points as in Prospect theory, we propose that the explanation for the participants' different preferences is related to the opposite evaluative polarities of the antonymic words used to communicate the outcomes of the options to the participants, i.e. *save* (positive) and *die* (negative) and the effect that negation has when it combines with *save* and *die* as in '*no people will be saved*', i.e. all die, and '*nobody will die*', i.e. all saved.

The purpose of this study is to provide a model that can be used to make predictions about people's choices when they are asked to make risky choice decisions related to the ADP. We argue that it is the evaluative polarity of the words used to communicate the options that is the crucial factor in decision making, rather than the domains of gains and losses as argued in Prospect theory. The predictions for how the evaluative polarities of the expressions mentioned above determine people's choices are analyzed from the perspective of different interpretive models of the *Gains-and-Losses Account* and the *Evaluative-Polarity Account*. We explore if and to what extent words such as *save* or *die* may make participants risk seeking or risk averse. Our reasoning is presented step by step, starting with two versions of the ideas advanced in Prospect theory, namely the *Gains-and-Losses Account* (Section 2.1) and the *Minimal-Gains-and-Losses Account* (Section 2.2). Using those as stepping-stones, we then develop three new models for the analysis of lexical semantic interpretations of the ADP, where evaluative polarity words are crucial cues in decision making. We begin with a simple version of what we refer to as the *Minimal-Evaluative-Polarity Account* (Section 3.2), which is further developed into the *Relative-Evaluative-Polarity Account* (Section 3.3) and even further into the *Extended-Relative-Evaluative-Polarity Account* (Section 3.4). The *Relative-Evaluative-Polarity Account* and the *Extended-Relative-Evaluative-Polarity Account* are the versions that differ most radically from the accounts based on Prospect theory, described in Sections 2.1 and 2.2. Their predictive strengths are empirically tested in two different experiments. The results of the experiments are reported in their respective section (Sections 3.3.1 and 3.4.1). Section 4 summarizes the results of the whole study in a general discussion. Two main claims are made: The evaluative polarity expressed through the words in the ADP influence people's risky choice preferences, and evaluative polarity is more important than domains when it comes to understanding framing effects of the ADP type.

2. The Gains-and-Losses Accounts

As already mentioned in the introduction, the options that decision makers choose between in the ADP are perfectly symmetrical. This means that any theory that sets out to explain the framing effect by reference to gains or losses will have to explain why 200 people saved out of 600 expected to die, is a gain, whereas 400 dead out of 600 expected to die is a loss. We introduce two different ways in which Prospect theory can explain this finding, first within the Gains-and-Losses Account, and then within the slightly modified Minimal-Gains-and-Losses account. Since Prospect theory is a purely formal theory, a number of assumptions have to be added in order to explain the ADP. The two versions presented in this

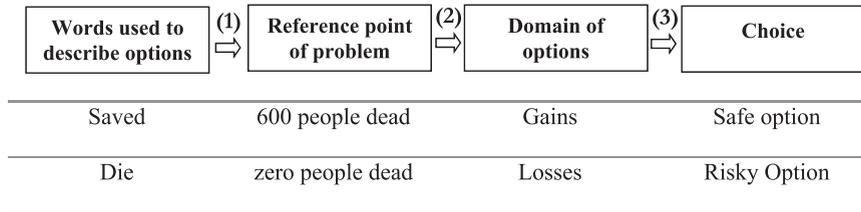


Fig. 1. The flowchart summarizes the Gains-and-Losses Account of the framing effect (for details on (1)–(3), see text). The table illustrates how the account works for Versions 1 (first row) and 2 (second row) of the ADP.

section represent our understanding of how this could be done. The discussion of the two Gains-and-Losses Accounts is used as a springboard and a motivation for the main objective of this study, namely to propose a linguistically oriented attempt to explain the framing effect of the ADP, which we do in Section 3.

2.1. The Gains-and-Losses Account

The Gains-and-Losses Account in this section is one out of potentially several representations of how the framing effect can be explained in Prospect theory. It is designed to clarify the relationship between reference points and domains. The Gains-and-Losses Account states that the words (*saved* or *die*) used to describe the options determine which reference point in the cover story (600 people dead or 0 people dead) is the reference point of the problem. The reference points, in turn, determine whether an option is seen as belonging to the domain of either gains or losses. A positive deviation from the reference point is a gain, and a negative deviation a loss (Tversky and Kahneman, 1981:454). Once the option is appropriately fixed as either a gain or a loss, Prospect theory predicts that participants are more likely to be either risk averse or risk seeking (Tversky and Kahneman, 1981:453). The Gains-and-Losses account is based on three assumptions:

- (1) *The words used to describe options determine the reference point;*
- (2) *The reference point defines the domain of the option (gain or loss);*
- (3) *The domain influences the choice.*

The relation between Assumptions (1), (2) and (3) is depicted by means of a flowchart in Fig. 1. The flowchart describes how the words used to describe options (1) determine the reference point (2) which in turn determines the domain (3), which influences the choice.

Starting from the end of the flow, we know that there is evidence in the literature that the domain influences the options, as stated in Assumption (3). This is a result of the so-called reflection effect, which shows that participants behave differently when they face gains and losses, i.e. in cases where the expected outcomes are *not* equivalent (see Markowitz, 1959; Kahneman and Tversky, 1979 for early empirical evidence; note, however that Fagley, 1993 claims that the evidence is mixed). This is evidence that goes beyond that which is provided by the ADP, which is of importance for the explanatory power of the model. There is not much that can be added to Assumption (2), which comprises the definition of domains in terms of reference points according to Tversky and Kahneman (1981).

Assumption (1) predicts that the words used to describe the options determine the reference point. This assumption is problematic because it is by no means a trivial operation to infer people's reference points from their choices (Fischhoff, 1983). An important issue for the Gains-and-Losses Account is how the reference point changes with the words used to describe the options. The Gains-and-Losses Account provides no such explanation. It simply assumes that it does, as pointed out by, among others, Fujii and Takemura (2003:54). Note also in relation to Assumption (1) that negation is not taken into account. The words referred to in this assumption are only *saved* and *die*, and not the contributions of negation in combination with *saved* and *die*. The model has nothing to say about the meaning of *not saved*, and *not die* or how possible reversals of evaluative polarity influence the participants' decisions (cf. Kühberger, 1995:231, and the discussion below).

One explanation for how reference points may become fixed is provided if we first assume that Version 1 of the ADP (the version using *saved*) activates the reference point 600 people dead, while Version 2 (the version using *die*) activates the reference point 0 people dead. If this is the case, then 200 saved out of 600 expected deaths belongs to the domain of gains at the same time as 400 dead out of 0 dead belongs to the domain of losses.¹ Thus, 200 saved out of 600 expected

¹ A natural extension of this definition is to classify the option as belonging to the domain of gains, if its expected value is a positive deviation from the reference point, and the reverse for losses.

to die counts as a positive deviation from the reference point (a gain), while 400 dead out of 0 possible non-deaths is a negative deviation from the reference point (a loss). Different versions of this explanation can be found in Maule (1989:165), Kühberger (1995:231) and McKenzie and Nelson (2003:596). In our comment on Assumption (1) above, we only touched upon the most probable mechanism, not *how* this happens. Why would the words used to describe options change the reference point? A promising route is to appeal to the notion of *information leakage* as proposed by Sher and McKenzie (2006, 2008, see also McKenzie and Nelson, 2003) and the subsequent development of their idea by Holleman et al. (2009). Sher and McKenzie (2006, 2008) argue that the communicative frames evoked by the words used in a given situation leak information. For instance, if a glass is described as *now being half full*, it is natural to infer that the glass was empty before, in which case the reference point would be a-less-than half full glass or a completely empty glass. By analogy, when it is said in the ADP that 200 people will be saved if a certain rescue program is implemented, it may be inferred that nobody will be saved if the program is not implemented, with 600 people dead as the reference point. The same reasoning could be used to argue that the reference point equals zero people dead when *die* is used because the point of reference would be that it was inferred that they all survive. Sher and McKenzie (2008) claim that participants that prefer Program A (the safe option) are more likely to frame this program as a gain, using *saved*. This indicates that information leakage can occur in the ADP and that participants are sensitive to this aspect.

It should be noted that Sher and McKenzie (2006, 2008) have investigated information leakage mainly in relation to attribute framing. This type of framing task involves the evaluation of a single option. In contrast, risky choice framing involves several options where different risks are compared (for a taxonomy of framing tasks, see Levin et al., 1998). Unfortunately, it is not a straightforward task to translate single-option-evaluation in attribute framing into several-option risky choice framing. Holleman et al. (2009) also point out that the experiments carried out by Sher and McKenzie (2006) do not take the communicative frame into account. Their proposal is instead an even more reflexive view based on how inferences are made to bridge the gap between psychological states and linguistic choices. Nevertheless, the idea of information leakage does provide us with a potential explanation for why evaluative polarity affects choices. We will come back to this idea also in the Evaluative-Polarity Accounts of the ADP.

2.2. The Minimal-Gains-and-Losses Account

For the sake of our argument in Section 3, this section presents a highly simplified variant of the Gains-and-Losses Account, in which words used to describe the options are allowed to directly determine whether the domain is gains or losses. In accordance with this interpretation, any option described by *saved* is a gain, and any option described by *die* is a loss. Consequently, the reference point has no influence on participants' decisions. The Minimal-Gains-and-Losses Account thus modifies Assumption (1) described in Fig. 1 into an Assumption (1_M). Together with Assumption (3) it explains the framing effect in the way depicted in Fig. 2.

(1_M) *The evaluative polarity of words used to describe the options directly determines the domain of the options so that positive evaluative polarity words set the domain to gains and words with a negative evaluative polarity set the domain to losses.*

Assumption (1_M) retains some of the features of Assumption (1), but it does not take negation into account, e.g. 'not saved'.

In Fig. 2 we see that the evaluative word meanings used to describe the options (1_M) directly determine the domain, and as a consequence also the participants' risk seeking behavior (Assumption 3). The question then is why these words have such a strong effect on participants' choices. In order to answer that question, it seems reasonable to appeal to the notion of information leakage, which should provide us with a mechanism for how evaluative words affect choices. Yet, it does not help us understand the reasons for why evaluative polarity is effective. What is clear though is the importance of

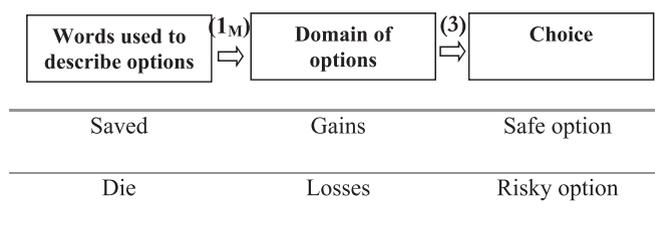


Fig. 2. The flowchart summarizes the Minimal Gains-and-Losses Account of the framing effect (for details see text). The table illustrates how the account works for Versions 1 (first row) and 2 (second row) of the ADP.

evaluative polarity in both Gains-and-Losses and Minimal Gains-and-Losses. It is therefore not satisfactory that neither negation nor communicative frames or differences in degrees of evaluative polarity have been discussed in the literature (with some notable exceptions). This is what motivated us to develop the Evaluative-Polarity Accounts described in Section 3 and discussed in Section 4.

3. The Evaluative-Polarity Accounts

In this section, we present our stepwise development of the Evaluative-Polarity Account of the ADP, starting with a minimal version that is subsequently extended and elaborated into the Relative-Evaluative-Polarity Account and the Extended-Relative-Evaluative-Polarity Account. Through this, we aim to make headway toward a more linguistically sophisticated approach to framing effects in the ADP, which focuses on the evaluative polarity of the words used to describe the options. Section 3.1 offers some linguistics preliminaries. It discusses the notion of frame in relation to the construal of opposition as expressed through *die* and *saved* in the different versions of the ADP. In Section 3.2, we describe the increasingly more elaborate steps of our Evaluative-Polarity Account, starting with the minimal version of the Evaluative-Polarity Account, which is similar to the Minimal-Gains-and-Losses Account, except for the fact that the latter also allows reference points to influence choice. In the Minimal Evaluative-Polarity Account (Section 3.2), we do not take negation into account at all. In Section 3.3, we present a more complex Evaluative-Polarity account, the Relative-Evaluative-Polarity Account, and empirically test whether negation matters in the ADP. The outcome of Experiment 1 prompts further development of the Evaluative-Polarity Account, in which frames, in the linguistic sense, are taken into account. This last step is the Extended-Relative-Evaluative-Polarity Account, which also allows for words to differ in their evaluative polarity.

3.1. Frames and polarity

As has already been mentioned, it is our contention that there are two aspects of the ADP that have to be taken more seriously to achieve a better understanding of the framing effect and the ADP. First, we have to acknowledge the role of frames – taken in their linguistic sense, which is broader than usually acknowledged in the decision making literature on framing effects – and second, we have to take negation, and how it influences evaluative polarity into account. The notion of frame in linguistics is concerned with the organization of lexical concepts of different scenes. Frame Semantics is a model of language use and understanding, within which encyclopedic knowledge, through the notion of frame, pervades all aspects of language use. Human beings cannot understand meanings of words without access to all the essential knowledge about the world and all essential knowledge of and about words and their use in different constructions, different frames and different discourses (Fillmore, 2006; Paradis 2005, 2015).

Applied to the ADP, we identify three frames incorporated in one another: The presentation of the cover story itself, the presentation of the rescue programs and the antonymic construction construals with *die* and *saved*. The most general level is the frame of the cover story, which in the case of the ADP is a DISASTER frame. The description of the national-level preparation of the handling of a serious disease evokes imminent danger, highlighted by lexical items such as *outbreak*, *unusual Asian disease*, *kill 600 people*. The next level consists of the two versions of the RESCUE frame. The rescue programs are contingent on the scene set through the cover story's DISASTER frame. Although the outcomes of the two versions are identical; the ways in which they are presented differ. The differences are construed through the evaluative polarity items, i.e. *die* and *saved*, with and without negation, with and without an explicit mention of the opposite pole. This is the EXISTENCE frame, expressed through *die*–*saved* and through negation, such as in *no people will be saved* and *no people will die*. The bipolar, antonymic EXISTENCE frame and the evaluative polarity of the words used to describe the options are crucial for how the rescue programs are interpreted and assessed.

In contrast to the Gains-and-Losses and the Minimal-Gains-and-Losses Accounts, negation in the Evaluative-Polarity Account cannot be ignored. It is well known that negation in text is not always used to reverse the truth of a proposition (e.g. Tottie and Paradis, 1982; Tottie, 1991; Giora et al., 2005; Giora, 2006; Paradis and Willners, 2006, 2013; Giora et al., 2007, 2009, 2010; Horn, 2010). Instead it has important discursive functions. Negation can be used to emphasize a meaning range along a dimension in contrast to another range along the same dimension. For instance, *this is not good* may be used instead of *this is bad*, or *this is not bad* instead of *this is good*. Negation may also be used as an approximator or attenuator, e.g. *not strong* may be used to attenuate the meaning of *strong* and instead evoke a meaning similar to *fairly weak*.

In our case, the negated contexts in the ADP are *no people will be saved* and *nobody will die*. In contrast to meanings such as 'bad' and 'good', which are scalar, 'saved' and 'die' are absolute. They do not indicate ranges on an unbounded scale, but are opposite properties on either side of a definite boundary, i.e. people either die or are saved (Paradis, 2001; Paradis and Willners, 2011; Jones et al., 2012). This absolute, bounded, distinction between *die* and *saved* is important for the equivalence between the first option of Version 1 and the first option of Version 2 of the ADP. Using judgment experiments, Paradis and Willners (2006, 2013) show that bounded antonymic meanings are interpreted as opposites of

their negated bounded antonyms. This then means that when *saved* and *die* are in the scope of negation they express opposite properties. *No people saved* in the context of the ADP means ‘all die’, and *nobody will die* means ‘all are saved’. In the former, a positive word meaning, *saved*, is negated, while in the latter case a negative word meaning, *die*, is negated. This fact may have mitigatory consequences.

Like us, Mandel (2014) questions the assumption of the ADP that language users’ understandings of options are equivalent. While we question the assumption that the various different evaluative expressions with *die* and *save* are interpreted in the same way, he raises another point that underlies the formulation of the ADP, namely that the numerical expressions, ‘200’ and ‘1/3’, are in fact interpreted by people as completely equivalent values. He correctly points out that in the linguistics literature on quantifiers there is general consensus that interpretations of numerical expressions are sensitive to the contextual factors. It is not the case that numerical expressions are always understood as exact values. On the contrary, it is well known that language users’ interpretations of quantifiers are predictably shaped by the context in which they occur. Language users may interpret meanings of numerical expressions such as the ones used in ADP either as ‘exactly’, ‘at least’ or ‘at most’ (e.g. Horn, 1989). Mandel argues that the participants, in fact, made linguistically rational and predicable choices. Either they chose the option that maximized the outcome (‘lives saved’) or their choices were consistent across the frames when the options were of equal value. Although there has been some debate as to the validity of this conclusion (e.g. see Chick et al., 2016, and below), we agree that speakers’ understanding of natural language use, irrespective of whether the expressions are numbers or words, has to be taken seriously.

It deserves to be pointed out that evaluative polarity has previously been discussed in connection with the framing effect. For instance, Yamagishi (1996, 2002) discusses evaluative polarity as a measure of *intrinsic pleasantness* and shows that it moderates certain aspects of the framing effect. Our concept of evaluative polarity also has to do with the *pleasantness of the whole decision problem*. It is related to what is often called the frame (Kühberger, 1995) or descriptor frame (Mandel, 2001). There are also previous studies of the effects of evaluative polarity. For instance, Kessler et al. (1996) point out that the value of an outcome is not only caused by gains or losses, but also by the value of the objects that could be gained or lost. In their experiments, participants were asked whether they prefer the risky choice of 85% chance to gain 10 minutes of pain, or a safe gain of 8 minutes of pain (Kessler et al., 1996:255). The majority of the participants chose the risky option. It is worth noting, however, that evaluative polarity here refers to the objects of choice, not to the words used to describe the options. Another example is Mandel (2001:64), who argues that there are two kinds of gains-and-losses descriptions, one of which he refers to as the descriptor frame. In the ADP, the descriptor frame refers to the use of either *saved* or *die*. Thus the descriptor frame of a problem determines what we refer to as the evaluative polarity of the problem. Mandel (2001:65) found that when *saved* was used there was more risk aversion than when *die* was used. In this section, we present a series of Evaluative-Polarity Accounts of the ADP in order to make a contribution to the important discussion in this field.

3.2. The Minimal-Evaluative-Polarity Account

The simplest way in which the evaluative polarity of words can influence participants’ behavior is by letting evaluative polarity directly determine risk-seeking behavior as in the Minimal-Gains-and-Losses Account. For this reason, we start our exposition of the Evaluative-Polarity Accounts with a Minimal-Evaluative-Polarity Account that allows the evaluative polarity of the words used to describe the options to influence risk-seeking and risk aversion directly, at the same time as it also allows the reference point to affect the choice. In order to arrive at a coherent account we cannot allow the reference point to shift with the words used to describe options (in that case we would not need reference points). Instead the Minimal-Evaluative-Polarity Account assumes one fixed reference point through all versions. We take the reference point to be the cover story’s expected outcome, i.e. 600 people dead in the original version of the ADP. This means that any reference point will fit this model as long as it remains fixed.

One of the fundamentals of the Evaluative-Polarity Account is that it allows for the words used to describe the options to define the evaluative polarity of the problem. In Version 1 of the ADP, for instance, the options are described by *saved*, which evokes a bounded (definite), non-scalar, positive meaning. This version of the ADP thus has positive evaluative polarity, i.e. the *problem* itself has positive evaluative polarity. Similarly, Version 2 of the ADP, where the options are described in terms of *die*, evokes a bounded, non-scalar, negative meaning and the problem has negative evaluative polarity. When a problem has positive evaluative polarity, people are expected to choose the safe option and are thus risk averse, and vice versa for negative evaluative polarity. In this minimal version of the Evaluative-Polarity Account, we also ignore negation. We return to the issue of negation in the versions of the Evaluative-Polarity Accounts described below. The Minimal-Evaluative-Polarity Account is best described by adding two assumptions to Assumptions 2 and 3 of the Gains-and-Losses Account.

- (4) *The words used to describe the options define the evaluative polarity of the problem;*
- (5) *The evaluative polarity of the problem influences the choice.*

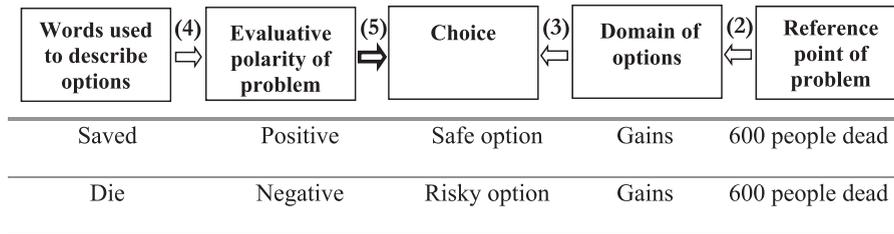


Fig. 3. The flowchart summarizes the Evaluative Polarity Account of the framing effect. The reference point is assumed to be fixed at the ADP cover story's expected outcome (for details on Assumptions 2–5, see text). The table illustrates how the Evaluative Polarity Account works for Versions 1 (first row) and 2 (second row) of the ADP. The arrow leading from evaluative polarity to choice is thicker than the arrow leading from domain to choice in order to show that evaluative polarity has a stronger influence on choice than domain (gains-and-losses) in the account.

The Minimal-Evaluative-Polarity Account in Fig. 3 differs from the Minimal-Gains-and-Losses account in Fig. 2 in that it allows both reference points and the evaluative polarities to influence the choice. It differs from the Gains-and-Losses Account in Fig. 1 in that decisions are influenced by both the domain of options (Assumption 3) and by the evaluative polarity of the problem (Assumption 5). More specifically, in the Evaluative-Polarity Accounts, it is also hypothesized that the effect of the evaluative polarity (Assumption 5) is stronger than the effect of the domain (Assumption 3), which is why the arrow leading from evaluative polarity to choice is thicker than the arrow leading from domain to choice. This hypothesis is necessary under the assumption of a fixed reference point, otherwise the Evaluative-Polarity Accounts could not explain the results of the original ADP.

3.3. The Relative-Evaluative-Polarity Account

In the first Minimal-Evaluative-Polarity Account, like in the Gains-and-Losses, and Minimal-Gains-and-Losses Account, negation is ignored, i.e. *2/3 probability that no people will be saved* in Version 1 and *1/3 probability that nobody will die* in Version 2. We have already pointed out that this is problematic since negated antonyms might not be completely synonymous with their non-negated antonyms. We acknowledge the findings in Paradis and Willners (2006, 2013) that negated bounded antonymic meanings, such as *dead* and *alive*, are indeed interpreted as synonyms of their negated counterparts, i.e. *not dead* is interpreted in the same way as *alive* and vice versa. Like them, we also make use of negated and non-negated antonyms, but, there is a difference between their work and ours in that their experiments are judgement experiments that tap into the interpretations of antonymic words in isolation. In our study, the stimuli are presented in a very specific context and the negated and non-negated antonymic constructions are examined in relation to one another as well as to an evaluative scale ranging from negative to positive.

In the Relative-Evaluative-Polarity Account we introduce the evaluative polarity of an *option*. This account thus deviates from Assumptions (4) and (5) of the first Minimal-Evaluative-Polarity Account. According to Assumption (4), the evaluative polarity of Version 1 of the ADP is positive because the meaning of *saved* is positive. This raises the question of whether the evaluative polarity of both the options of this problem is positive. We argue that the evaluative polarity is positive for the safe option, i.e. *200 people are saved*, but not for the risky option, i.e. *1/3 probability that 600 people will be saved* and *2/3 probability that no people will be saved*. The risky option has *mixed* evaluative polarity. It is described by *saved*, which evokes a positive meaning, and by the expression *no people saved*, which evokes a negative meaning in spite of the fact that the positive word *saved* is used. The idea that an option may be associated with evaluative polarity is related to what Kühberger calls a frame (1995:231, Table 1). According to him, the frame of the safe option is positive, while the frame of the risky option is mixed since it contains expressions with negation, e.g. *no people saved*. Note that the Relative Evaluative-Polarity Account is very similar to Fuzzy Trace Theory (Reyna and Brainerd, 1991; Kühberger and Tanner, 2010; Chick et al., 2016; Reyna et al., 2014). Here, the notion of 'gist' is used to capture that participants do not attend to the full information available, but instead focus on the lowest level representation. In this case, this involves saving some lives for certain or saving some or no lives with a probability (Chick et al., 2016:240). According to Fuzzy Trace Theory, participants, in the positive version of the ADP are said to choose between saving *some* lives and for the risky option between saving *some* or *none*. We want to point out that although Fuzzy Trace Theory acknowledges the importance of negation, it does not expand on the evaluative polarity of options. It can therefore not accommodate the Extended Relative Evaluative Polarity account, which we present in Section 3.4. Mandel (2001:64, Table 3) also uses the term *outcome frame* to refer to the meaning of expressions that include negation, such as *no people saved*. He does not, however, consider decision options with mixed evaluative polarity. This factor, however, is of vital importance for our understanding of the connection to the meaning of the evaluative polarity words in the Relative-Evaluative-Polarity Account.

Table 1

A summary of the predictions of the Relative Evaluative Polarity Account (see Fig. 4). See text for descriptions of the ADP cover story and of the ADP Versions 1, 2, 3, and 4.

| Range of outcomes | Description of safe option | Evaluative polarity of safe option | Description of risky option | Evaluative polarity of risky option | Relative evaluative polarity (predicted choice) |
|-------------------|---|------------------------------------|---|-------------------------------------|---|
| 0 dead | 200 saved (Version 1) | Positive | 1/3 that 600 saved and 2/3 that 0 saved (Version 1) | Mixed | Safe option |
| | 200 saved and 400 not saved (Version 3) | Mixed | 1/3 that 600 saved (Version 3) | Positive | Risky option |
| | 400 die (Version 2) | Negative | 1/3 that 0 die and 2/3 that 600 die (Version 2) | Mixed | Risky option |
| | 400 die and 200 not die (Version 4) | Mixed | 2/3 that 600 die (Version 4) | Negative | Safe option |
| 600 dead | Reference point introduced by cover story | | | | |

The analogues of Assumptions (4) and (5) for the Relative-Evaluative-Polarity Account are as follows:

- (6) *The words used to describe the options define the evaluative polarity of the options;*
- (7) *The relative evaluative polarity of the options influences the choice.*

Assumption (7) states that participants choose an option with positive evaluative polarity rather than an option with mixed evaluative polarity and an option with mixed evaluative polarity rather than an option with negative evaluative polarity. We have replaced Assumption (4) and Assumption (5) with (6) and (7) respectively. The Relative-Evaluative-Polarity Account of the framing effect with a fixed reference point is shown in Fig. 4.

In Fig. 4 we see that the words used to describe the options (Assumption 6) affect the evaluative polarity of the options (Assumption 7) and that this influences choice to a higher degree than the domain of the options (Assumption 3, given by the reference point, Assumption 2). The Relative-Evaluative-Polarity Account allows negation to influence participants' behavior in the ADP, whereas the Minimal-Evaluative-Polarity Account does not. The Relative-Evaluative Polarity Account predicts that the relative evaluative polarity of the option will exert stronger effects on participants' risk seeking behavior than the evaluative polarity of the problem. The Minimal-Evaluative-Polarity Account predicts the opposite since it ignores negation. The hypothesis of relative evaluative polarity is given in (H1).

H1. *The relative evaluative polarity of the option, i.e. that the safe option has higher evaluative polarity or the risky option has higher evaluative polarity, has stronger influence on the choice than the evaluative polarity of the problem, i.e. the positive or the negative meaning of the words.*

It is important to note that neither Gains-and-Losses nor Minimal Gains-and-Losses make any predictions to this effect, although it could, perhaps, be argued that the Minimal-Gains-and-Losses Account should make the same prediction as the Minimal-Evaluative-Polarity Account. This is because Prospect theory and our two derivative models of Prospect theory ignore linguistic aspects of the framing effect. Note also that a similar prediction to H1 can be derived from Fuzzy Trace Theory, but again since this theory is mute on the evaluative polarity of the options, it cannot move beyond negation in its explanation of the framing effect.

3.3.1. Experiment 1

3.3.1.1. *Procedure, participants and design.* All in all, 259 participants (average age = 22.6 years; standard deviation = 4.8 years) participated in Experiment 1. They were undergraduate students of philosophy and computer

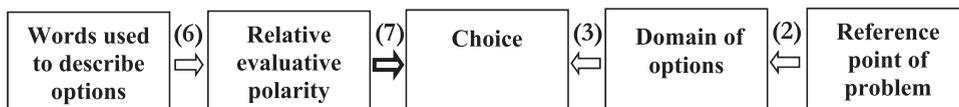


Fig. 4. A summary of the Relative Evaluative Polarity Account of the framing effect with a fixed reference point (for details on 2, 3, 6, and 7, see text). This account differs from Evaluative Polarity Account (Fig. 3) in that it specifies how the evaluative polarity of individual options – rather than the evaluative polarity of the whole problem – influences choice. The arrow leading from relative evaluative polarity to choice is drawn thicker than the arrow leading from domain to choice in order to show that in the Relative Evaluative Polarity Account the relative evaluative polarity of options has a stronger influence on choice than domain.

science at Lund University, all fluent in English. We used the cover story of the original ADP with the reference point of 600 people dead. After reading the story, the participants of each of the four groups were asked to choose between Programs A and B, described in different words in the four versions given below. Within each version the order of presentation of the programs was counterbalanced.

Two of the problems tested were the original Versions 1 and 2 of the ADP described in the introduction. The other two were new problems, namely Versions 3 and 4 described below. In Version 1 of the original ADP, as Kühberger (1995) points out, the safe option, *200 people will be saved*, is not fully described as it is not explicitly stated what will happen to the remaining 400 people. Is it the case that they will die or is it the case that they will never completely recover from the disease? On the other hand, the risky option is described in full in the sense that it is explicitly stated that there is 1/3 probability that everyone will be saved and there is 2/3 probability that everyone will die. Version 3 is a variant of Version 1. It has been changed so that the safe option is fully described, while the risky option is not. Similarly, Version 4 was obtained by changing Version 2 of the original ADP so that the safe option is fully described and the risky option is not:

Version 3

- If Program A is adopted, 200 people will be saved and 400 people will not be saved.
- If Program B is adopted, there is 1/3 probability that 600 people will be saved.

Version 4

- If Program A is adopted, 400 people will die and 200 people will not die.
- If Program B is adopted, there is 2/3 probability that 600 people will die.

Like the Minimal-Evaluative-Polarity Account, the Relative-Evaluative-Polarity Account predicts that the safe option will be chosen in Version 1 and the risky option will be chosen in Version 2. For Versions 3 and 4, however, the predictions of the Relative-Evaluative-Polarity Account are different from the predictions of the Minimal-Evaluative-Polarity Account. In Version 3 first, the safe option has mixed evaluative polarity and the risky option has positive evaluative polarity. Thus the Relative-Evaluative-Polarity Account predicts that the risky option will be chosen. The account based on the Minimal-Evaluative-Polarity Account predicts that the safe option will be chosen. In Version 4, the safe option has mixed evaluative polarity and the risky option has negative evaluative polarity. The Relative-Evaluative-Polarity Account predicts that the safe option will be chosen, while the Minimal-Evaluative-Polarity Account predicts that the risky option will be chosen. Table 1 summarizes the predictions of the Relative-Evaluative-Polarity Account for the four versions.

Table 1 shows that our design has two factors: The evaluative polarity of problem and the relative evaluative polarity of option. The evaluative polarity of the problem has two levels: It is positive when the word used is *saved* and negative when word used is *die*. The relative evaluative polarity also has two levels that indicate which option has the higher evaluative polarity: The safe option or risky option.

3.3.1.2. *Results.* The results are shown in Fig. 5. In Version 1, 39% of the participants chose the risky option and in Version 2, 67% of the participants chose the risky option. This replicates the results in the original ADP and is consistent

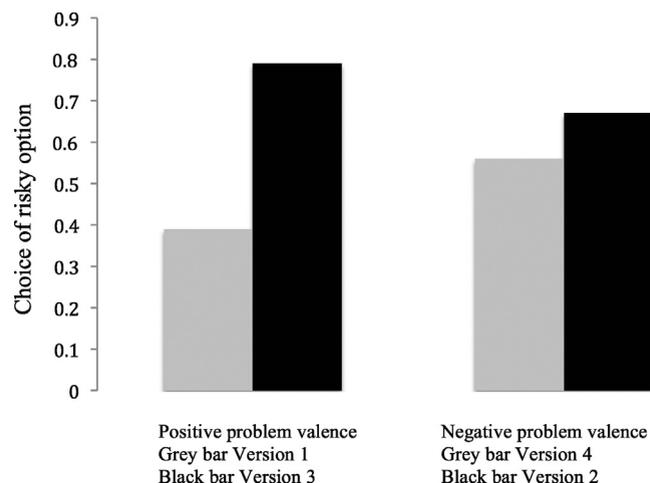


Fig. 5. Experiment 1: The influence of the relative evaluative polarity of the options on the choice is stronger than the influence of the evaluative polarity of the problem. Gray bars represent Versions where the relative evaluative polarity is higher for the safe option (Versions 1 and 3). Black bars represent Versions where the relative evaluative polarity is higher for the risky option (Versions 2 and 4).

Table 2

A summary of the logistic regression results for the contrasts between Versions 1–3 and 2–4 (the Relative Evaluative Polarity Account predicts differences) and between Versions 2–3 and 1–4 (the Minimal Evaluative Polarity Account predicts differences).

| Contrast | Coefficients | Standard errors | Z-values | P-values |
|---------------------|--------------|-----------------|----------|----------|
| Version 1–Version 3 | 1.76 | 0.39 | 4.49 | <0.0001 |
| Version 2–Version 4 | –0.47 | 0.37 | –1.28 | 0.20 |
| Version 2–Version 3 | 0.64 | 0.40 | 1.59 | 0.11 |
| Version 1–Version 4 | 0.65 | 0.36 | 1.83 | 0.07 |

with all accounts in Figs. 1, 2 and 4.² In Version 3, 79% of the participants chose the risky option, while in Version 4, 56% of the participants chose the risky option. Thus the Relative-Evaluative-Polarity Account was partly supported.

We performed a logistic regression analysis, and contrast-coded the four versions, so that the estimates of the analysis represented the contrasts between Version 1 and 3, and between Version 2 and 4, which are the contrasts relevant for the Relative-Evaluative-Polarity Account, and then once more between Version 2 and 3; and between Version 1 and 4, which are the contrasts relevant for the Evaluative-Polarity Accounts. Table 2 shows the coefficients, standard errors, z-values and p-values for each contrast, i.e. between Version 1 and 3, Version 2 and 4, and between Version 2 and 3; Version 1 and 4. The contrast between Versions 1 and 3, which is relevant for the Relative-Evaluative-Polarity Account, is significant at $p < 0.0001$. The one between Versions 1 and 4, which is relevant for evaluative polarity, has a $p = 0.07$. Neither of the two contrasts including Version 2 is significant. Hypothesis (H1) is therefore only partly supported.

3.3.1.3. Discussion. The relative evaluative polarity of the options affects people's decisions in this extended ADP. The influence of relative polarity of the options is in some cases stronger than that of the evaluative polarity of the problem, but not always. When options are described using the word *save*, i.e. positive evaluative polarity words, negation appears to matter in the way proposed by the Relative-Evaluative-Polarity Account. The importance of negation has also been shown in previous research, in particular by Kühberger (1995). However, we observe a differential effect of relative-option-evaluative polarity for Versions 3 and 4. We will return to this below, and in the Extended-Evaluative-Polarity Account, because the differential effect appears to be best explained by differences in how negation affects the evaluative polarity of options.

It deserves to be mentioned that there are some problems with the interpretation of Experiment 1. In particular, Program B of Version 3, *1/3 probability that 600 will be saved*, and Version 4, *2/3 probability that 600 die*, are only equivalent if we assume that a statement saying that a certain proportion of the inhabitants are saved or die entails that the rest are not saved or will not die. If participants instead interpret the statement so that the un-mentioned proportion is 'not-all saved', the outcomes of Version 3 and 4 will differ. To some extent the same criticism applies to Program A in Versions 1 and 2 of the original ADP too, where it is not explicitly stated what happens to the individuals that are not saved, or do not die (cf. Mandel, 2014). It is, however, possible to argue that the contrast is clearer when probabilities are not included. It is difficult to control for this factor, since a further specification of Program B would lead to a change in the relative evaluative polarity of the problem. When Chick et al. (2016) disambiguated the interpretation in the ADP and variants of it, using either lives or monetary outcomes, the framing effect remained robust. The participants received disambiguating instructions prior to responding to the various framing problems, for instance "... if we specify that there is a probability that everyone will be saved, if that does not happen, no one will be saved. Similarly, if we specify that there is a probability that no one will be saved, if that does not happen, everyone will be saved" (ibid: 254). An interesting extension for further research would be to examine the evaluative interpretation of the probability statements in themselves, perhaps in the manner indicated by Fuzzy Trace Theory. Potentially statements such as '1/3 probability' may give rise to evaluative polarity in themselves.

Note also that Kühberger (1995) presented participants with fully specified versions of the outcomes for both the risky and the safe options so that Program A, for instance, was described in full as *200 will be saved and 600 will not be saved*, and as Program B is in the original Version 1 of the ADP *1/3 probability that 600 will be saved and 2/3 probability that no people will be saved*. With this description, 62% of the participants in Kühberger's Experiment 1 are risk seeking in the version using positive evaluative polarity words, and for the equivalent, but negative, 57% are risk seeking. Kühberger's replication of the standard ADP yielded a clear (but not as strong as in the original experiments) framing effect with 48% of

² It also confirms that the Swedish participants, responding to an English version of the ADP, behave in the same way as the native speakers did. There is no major distortion of the results due to the language used in the experimental trials, but see Oganian et al. (2016), for a detailed discussion of language switching and the framing effect.

the participants being risk seeking for the version with positive evaluative polarity words, and 78% risk seeking in the version with negative evaluative polarity words. The fact that the framing effect disappears when options are fully described, i.e. when all information is explicitly mentioned in the text, suggests that the evaluative polarity of the words used to describe the options are very important for participants' choices, and, in our opinion, it gives credibility to our interpretation of Versions 3 and 4.

Chick et al. (2016) also used a truncated versions of the ADP, but with only one of the risky options specified, i.e. “2/3 probability no one saved” and “1/3 probability 600 saved” respectively. They did not, however, fully specify the safe options as in Experiment 1, but rather removed parts of the information. In the case specified above, Fuzzy Trace Theory would predict an increased framing effect since the gist of the options is that it is certain that *some* people were saved compared to *none* were saved, and the attraction of the safe option is thus higher than in the original ADP where participants were asked to choose between that it is certain that *some* people were saved and *some* or *none* were saved. The Relative-Evaluative-Polarity Account makes the same prediction. The truncated versions of Chick, Reyna and Corbin, support both theories, as do the results of the positive evaluative polarity versions of ADP in Experiment 1. Note, however, that Fuzzy Trace Theory uses a different mechanism to make these predictions, namely memory traces rather than evaluative polarity. As we will see below, evaluative polarity allows us to post hoc explain the differential outcomes of the positive and negative evaluative polarity options (e.g. *save* vs. *die*).

In the Relative-Evaluative-Polarity Account and in Experiment 1, the evaluative polarity of an option is determined by a simple algebra on the basis of the evaluative polarity evoked by the words used to describe the options, i.e. the evaluative polarity of the antonyms with and without negation. This is what leads to the similarity to Fuzzy Trace Theory. For example, the evaluative polarity of the option *200 people will be saved* and *400 people will not be saved* is assumed to equal the value of *saved* plus the evaluative polarity of NEGATOR + *saved*. This assumption has received experimental support showing that the participants' interpretations of bounded antonymic meanings are interpreted as one another's absolute opposites, i.e. *die* is interpreted as a synonym of *not alive* and *not alive* as a synonym of *die* (Paradis and Willners, 2006, 2013). Now, it goes without saying that the algebra is a simplification. First, it implies that the number of the people that will be saved or that will die is irrelevant, which is a very strong assumption. A second implication is that the sum of the evaluative polarity of a word and of the value of its negation is constant across words. For example, the evaluative polarity of the option *200 people will be saved* and *400 people will not be saved* is assumed to equal the evaluative polarity of the option *400 people will die* and *200 people will not die* – both have mixed evaluative polarity (see Table 1). It seems to be the case that this is not true because, if it were true, we would not have observed a differential effect of the relative-option-evaluative polarity for the evaluative polarity conditions of Versions 3 and 4. The effect of the relative evaluative polarity of the problem is significant when the problem evaluative polarity is positive but not when it is negative. This result can be explained on the basis of the assumption that the value of NEGATOR + *die* is smaller than the absolute value of NEGATOR + *saved*. In order to make sure that this is the case, we asked 124 undergraduates at Lund University to rate the two words in terms of their pleasantness on a scale from –5 to +5. The average rating of *not die* was +0.7, while the average rating of *not be saved* was –2.8. The words were embedded in a list of 14 near synonyms such as *survive* and *lose one's life* and the order of the words was counterbalanced across the participants. According to a *t*-test, the difference between the absolute values was significant at the 0.001 level. This seems reasonable since *not be saved* implies that someone missed the chance of being saved – a clearly unpleasant outcome – while *not die* is a description of survival, in which case the implication is that *status quo* is retained. Note that this explanation is not available in Fuzzy Trace Theory.

A slightly different version of the algebraic model is also used in the work of Kühberger (1995:231, Table 1). In his work, the algebra is applied to the frame describing the options and not to the evaluative polarity of an option. Thus, the option, *200 people will be saved* and *400 people will not be saved*, has a mixed evaluative polarity in our study whereas for Kühberger (1995), it has a mixed frame. It may be argued that, in the end, the concepts of evaluative polarity, gist, and frame have similar theoretical functions. One difference that may be important for future theorizing is that frames have not been related to ‘intrinsic pleasantness’, whereas evaluative polarity has (Yamagishi, 2002). This is an important theoretical difference since Kühberger's and Fuzzy Trace Theory's focus is on whether the options are fully described or not, i.e. whether and how they are specified, and what will happen to those that do not die or are not saved. This account cannot accommodate possible variation in the values of each evaluative word such as *not saved* vs. *die*. Therefore, Kühberger cannot explain the differences between Versions 3 and 4 in the way that we can through our Relative-Evaluative-Polarity Account.

3.4. The Extended-Relative-Evaluative-Polarity Account

The Extended-Relative-Evaluative-Polarity Account uses an algebraic model that can take value differences in evaluative polarity into account. If participants directly respond to the valence, i.e. ‘the intrinsic pleasantness’, of the words used to describe the options (so far only *saved* and *die* vs. *not saved* and *not die*), the degree of evaluative polarity could

be used to predict the degree to which participants will be risk seeking or risk averse in different formulations of the ADP. The difference between Version 3 and Version 4 above is a case in point. In order to investigate this possibility, we propose a tentative Extended-Relative-Evaluative-Polarity Account, which adds the *valence*, where valence is the degree of evaluative polarity of each option. The Extended-Relative-Evaluative-Polarity Account predicts that the tendency to be risk seeking or risk averse is influenced by this graded valence. In this way, we can begin to address the effect that variation in valence has on the ADP.

In order to formulate the Extended-Relative-Evaluative-Polarity Account, we need to specify how the valence of the options is computed. We take the original version of the ADP as an example. The words that are of importance for the evaluative polarity algebraic model are in italics.

Version 1

- If Program A is adopted, 200 people will be *saved*.
- If Program B is adopted, there is 1/3 probability that 600 people will be *saved*, and 2/3 probability that *no people will be saved*.

Version 2

- If Program A is adopted, 400 people will *die*.
- If Program B is adopted, there is 1/3 probability that *nobody will die*, and 2/3 probability that *600 people will die*.

There are two things that have to be captured by the algebraic model. First, we need to retain the predictions of the Relative-Evaluative-Polarity Account, so that, in general, an option with positive evaluative polarity is chosen over one with mixed evaluative polarity, which in turn should be preferred to options with negative evaluative polarity. Second, we need to take the valence of the evaluative polarity words into account, and in particular the difference in valence between positive and negative evaluative polarity words that appear in the same option, such as Program B in the two versions above. In general, the larger the distance between the two evaluative polarity expressions, i.e. the distance between *saved*-NEGATOR + *saved*, and *die*-NEGATOR + *die*, the more contrastive strength the mixed option will have, in comparison to an unqualified, positive or negative evaluative polarity option. Both *saved* and *not saved* and *die* and *not die* are absolute opposites. Even though intermediate interpretations can be constructed, they may not be the most natural as is the case for meanings of expressions such as *good* and *not good*, or *bad* and *not bad* (Paradis, 2001; Paradis and Willners, 2006, 2013).

In addition, the algebra will have to account for versions of the ADP, where the fully specified options are those for Program A instead of for Program B in Experiment 1.

Version 3

- If Program A is adopted, 200 people will be *saved* and 400 people will *not be saved*.
- If Program B is adopted, there is 1/3 probability that 600 people will be *saved*.

Version 4

- If Program A is adopted, 400 people will *die* and 200 people will *not die*.
- If Program B is adopted, there is 2/3 probability that 600 people will *die*.

For the Extended Relative-Evaluative-Polarity Account, we therefore propose an algebraic model in two steps. First, the plus or minus sign for the difference in evaluative polarity between Program A and B is calculated based on the evaluative polarity of the programs with fully specified options and the distance between the evaluative polarity of the wordings. The mathematical sign is then used to predict which option participants will be most likely to choose. For Version 1, we determine the sign by taking the evaluative polarity of *saved* and subtract the distance between the evaluative polarity of *saved* and the evaluative polarity of NEGATOR + *saved*. For Version 2, it is obtained by taking the evaluative polarity of *die* and subtracting the distance between the evaluative polarity of *die* and the evaluative polarity of NEGATOR + *die*. For Version 3, instead, the sign is determined by taking the distance between the evaluative polarity of *saved* and the evaluative polarity of NEGATOR + *saved* and subtracting the evaluative polarity of *saved*, and, finally, for Version 4, the distance between the evaluative polarity of *die* and the evaluative polarity of NEGATOR + *die* and subtracting the evaluative polarity of *die*. Most likely, the sign will be positive for Program A in Version 1 and 4; hence we predict that the participants will choose this option. The sign for Program B in Version 2 and 3 will likely be negative; hence the participants will choose Program B. The second step of the model is achieved by using the distance between the words used to describe the option (*save*-NEGATOR + *save*, and *die*-NEGATOR + *die*) to determine the strength of the prediction. In this way, a participant with a large distance between the negated and non-negated option will most likely choose the program with a positive sign, because it will be the clearest positive outcome given that it has a larger numerical distance.

3.4.1. Experiment 2

In analogy with Hypothesis (H1), we state:

H2. The sign for the different options together with the distance between the evaluative polarity words used to describe options, i.e. the word and its negated variant, determines the degree of risk seeking, so that the option with a positive sign is more likely to be chosen the larger the distance between the evaluative polarity words used is.

3.4.1.1. Procedure, participants and design. All in all, 732 participants (average age = 22.9 years; standard deviation = 2.96 years) participated in Experiment 2. They were undergraduate students of social sciences, humanities and law at Lund University, all fluent in English

We constructed 12 different versions of the ADP using synonyms for the expressions used to describe the options of the opposite pairs: *gained–lost*, *survived–killed* and *lengthened–shortened* together with their negated variants. In the descriptions of the ADP problem, all the stimuli express evaluative polarity. Two of them are bounded, either-or antonym pairs and so are their negated variants, i.e. either lives were lost or not, either lives were gained or not, etc. One pair, *lengthen–shorten*, is gradable in the sense that lives can be shortened or lengthened, but the extent of the shortening or lengthening may be applied differently by different people. This word pair was included to also capture the effect of scalable meanings with opposite evaluative meanings. We used many versions in order to ensure sufficient variability in evaluative polarity. Each expression was used in two versions of the ADP; the original type where Program B is fully specified (as in Version 1 and 2 above), and the modified type where, instead, Program A is fully specified (as in Version 3 and 4 above). All versions can be found in Table 4, together with the assessments to each version by the participants.

After having responded to one particular version of the ADP (that is, only one out of the in total 12 possible versions), each participant was asked to rate the intrinsic pleasantness of each stimulus word and its negated variant on a scale from 1 to 11. The order of expressions was counterbalanced, and each participant responded to the ADP first, before rating the expression. (Note that we used 200 for all numerical expressions so that participants would focus on the valence of the word rather than the number). The mean valence of each expression is given in Table 3, which shows that there are clear differences between positive and negative statements, but, as expected, the magnitude differs across the stimulus pairs.

In order to be able to predict an outcome for the different versions of the ADP using the proposed algebra for the Extended Relative-Evaluative-Polarity Account, there has to be a difference between the evaluative polarity of Program A and B. For 17 of our participants this was not the case, and these were excluded from further analysis. Eleven additional participants were excluded for not having rated the words used in their particular version of the ADP. This left us with 704 participants all in all. For most ADPs, the sign of the difference between Programs A and B coincided with how the options would be characterized by the Relative-Evaluative-Polarity Account. Participants who did not display the differences predicted by Relative-Evaluative-Polarity were kept in the analysis. For the exact number of answers in the expected direction see Table 4. An analysis was performed for all individual participants and his or her individual rating of the stimulus expression, so that the evaluative polarities used to predict the likelihood of risk seeking behavior were unique for each participant.

3.4.1.2. Results. We performed a logistic regression analysis using the sign for the scenario (with a positive sign predicting risk averse and a negative sign predicting risk seeking behavior) and the distance between the expression used in the ADP and its negation as predictors. Table 5 displays the coefficients, standard errors, z-values and p-values for each. In particular we note that a larger distance in evaluative polarity between the negated and non-negated expressions, together with a positive sign predicts a lower number of risk seeking participants: -0.14 ($p < 0.01$). In Fig. 6, we chart participants with positive and negative signs and display how their likelihood of being risk seeking varies with the distance in evaluative polarity.

These results concern all versions of the ADP, i.e. 1–12. It might make sense also to zoom in on versions of the ADP that lie close to the original, and in which only the evaluative polarity words have been varied, that is the versions in which

Table 3
The mean valence (SE) of the expressions used in Experiment 2.

| Expression | Mean valence (SE) | Negated expression | Mean valence (SE) |
|---------------------------------------|-------------------|---|-------------------|
| 200 people will survive | 9.91 (1.31) | 200 people will not survive | 2.34 (1.40) |
| 200 people will be killed | 1.46 (0.74) | 200 people will not be killed | 8.14 (2.07) |
| 200 lives will be gained | 8.66 (1.71) | 200 lives will not be gained | 4.04 (1.64) |
| 200 lives will be lost | 2.33 (1.37) | 200 lives will not be lost | 8.07 (1.78) |
| 200 people's lives will be lengthened | 8.57 (1.50) | 200 peoples' lives will not be lengthened | 3.97 (1.63) |
| 200 people's lives will be shortened | 3.13 (1.38) | 200 people's lives will not be shortened | 7.60 (1.84) |

Table 4

The versions of the ADP used in Experiment 2 together with the number of participants that choose each option (in parenthesis).

| Version | | No. participants choosing option |
|------------|---|----------------------------------|
| Version 1 | If Program A is adopted, 200 people will survive. | 37 |
| | If Program B is adopted, there is 1/3 probability that 600 people will survive, and 2/3 probability that no people will survive. | 25 |
| Version 2 | If Program A is adopted, 400 people will be killed. | 15 |
| | If Program B is adopted, there is 1/3 probability that nobody will be killed, and 2/3 probability that 600 people will be killed. | 47 |
| Version 3 | If Program A is adopted, 200 people will survive and 400 people will not survive. | 25 |
| | If Program B is adopted, there is 1/3 probability that 600 people will survive. | 29 |
| Version 4 | If Program A is adopted, 400 people will be killed and 200 people will not be killed. | 31 |
| | If Program B is adopted, there is 2/3 probabilities that 600 people will be killed. | 28 |
| Version 5 | If Program A is adopted, 200 lives will be gained. | 44 |
| | If Program B is adopted, there is 1/3 probability that 600 lives will be gained, and 2/3 probability that no lives will be gained. | 17 |
| Version 6 | If Program A is adopted, 400 lives will be lost. | 15 |
| | If Program B is adopted, there is 1/3 probability that no lives will be lost, and 2/3 probability that 600 lives will be lost. | 47 |
| Version 7 | If Program A is adopted, 200 lives will be gained and 400 lives will not be gained. | 26 |
| | If Program B is adopted, there is 1/3 probability that 600 lives will be gained. | 35 |
| Version 8 | If Program A is adopted, 400 lives will be lost and 200 lives will not be lost. | 41 |
| | If Program B is adopted, there is 2/3 probabilities that 600 lives will be lost. | 17 |
| Version 9 | If Program A is adopted, 200 lives will be lengthened. | 32 |
| | If Program B is adopted, there is 1/3 probability that 600 lives will be lengthened, and 2/3 probability that no lives will be lengthened | 28 |
| Version 10 | If Program A is adopted, 400 lives will be shortened. | 25 |
| | If Program B is adopted, there is 1/3 probability that no lives will be shortened, and 2/3 probability that 600 lives will be shortened. | 38 |
| Version 11 | If Program A is adopted, 200 lives will be lengthened and 400 lives will not be lengthened. | 25 |
| | If Program B is adopted, there is 1/3 probability that 600 lives will be lengthened. | 42 |
| Version 12 | If Program A is adopted, 400 lives will be shortened and 200 lives will not be shortened. | 29 |
| | If Program B is adopted, there is 2/3 probability that 600 lives will be shortened. | 34 |

only Program B is fully specified. For this half of our participants we see an effect of distance in evaluative polarity between negated and non-negated expressions for versions with a positive sign, in this case, mostly versions where options are described using positively valenced evaluative polarity words such as *survive*, *gain* and *lengthened* (estimate 1.02, SE = 0.52, $p = 0.005$), but not for versions with a negative sign (estimate -0.45 , SE = 0.35, $p = 0.20$).

3.5. Discussion

Generally speaking, it seems to be the case that the distance between the evaluative polarity word used to describe the option and its negation can be used to predict the likelihood that a participant will be risk seeking, or risk averse, depending on the plus or the minus sign. The results appear to be sensitive to the type of evaluative polarity words that are used to describe the options, however. When we zoom in on ADPs with a traditional setup (in which only Program B is fully specified), we see that the effect only remains for (mostly) positively valenced evaluative polarity words.

This sensitivity might be because the differences across participants' assessments of the test items in the ADP are limited. The meanings of our test items are bounded, non-scalar meanings, such as 'saved', 'die' and so are their negated

Table 5

A summary of the logistic regression results when risk seeking is predicted using the *sign* (positive or negative) and the *distance* between the positive and negative evaluative polarity words used to describe the option.

| | Estimate | Standard errors | Z-values | P-values |
|-----------------|----------|-----------------|----------|----------|
| Intercept | -0.16 | 0.25 | -0.64 | 0.52 |
| Sign | 0.15 | 0.35 | 0.43 | 0.67 |
| Distance | 0.10 | 0.04 | 2.76 | 0.006 |
| Sign * Distance | -0.14 | 0.05 | -2.64 | 0.008 |

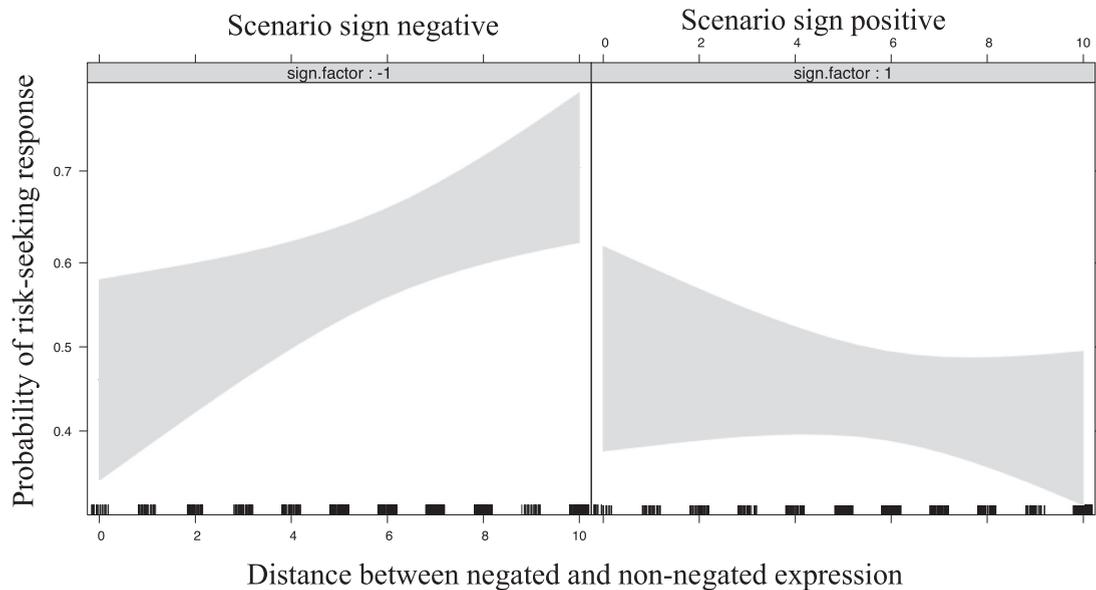


Fig. 6. The probability of risk seeking responses given distance in evaluative polarity between the negated and non-negated expressions used to describe the options (e.g. saved–NEGATOR + saved) for participants with a negative sign for the difference between options (left, risk seeking predicted) and a positive sign (right, risk aversion predicted) respectively.

versions. In fact, as has been mentioned before, when the participants were asked where on a scale they locate the meanings of *dead* and *alive*, they indicate that they occupy the position at the very ends of a given scale, close to the extreme boundaries. An interesting finding was that the positive results are stronger for the positive versions of the ADP than for the negative ones in both experiments (see Figs. 5 and 6). This is an intriguing finding, which we return in the Section 4, where we also compare and discuss other researchers' findings in relation to ours.

4. General discussion

The framing effect has been extensively studied in the literature. Due to both its theoretical and practical importance this is not surprising. It is theoretically important because it challenges the foundations of the dominant normative decision theory, one of the core assumptions of which is that choices are invariant, i.e. decision makers do not respond differently to options with identical outcomes. When this happens – as in the framing effect – the consequences are potentially dire. As mentioned in the introduction, the practical implications of the framing effect are perhaps even more important since it is known to affect people's behavior in real world decisions in different areas ranging from medicine to finance. In spite of these important and concrete applications, there is at present no theory that can explain why and when the effect occurs, as noted by Maule and Villejoubert (2007:25). We have approached the framing effect, as it occurs in risky choice, in a cross-disciplinary study combining insights from philosophy and psychology with research in linguistics. Our goal has been to provide an alternative explanation for the ADP, which does not rest on concepts such as *domain* (gains or losses). Instead, we have focused on the evaluative polarity of the words used to describe the decision options. The hypotheses tested were that (i) the words used to describe the options define the evaluative polarity of the problem, and (ii) the evaluative polarity of the problem (positive or negative) influences the choice. Both of them were borne out to some extent.

We developed a tentative model for how the evaluative polarity of options, as opposed to the domain of the problems, affects decision makers (Fig. 4). This model does not only focus on whether the words used to describe the problem have positive or negative evaluative polarity (*saved* or *die*) but also on the presence of negation in the description of the options. This is similar to variants of Fuzzy Trace Theory (e.g. Reyna and Brainerd, 1991; Kühberger and Tanner, 2010; Chick et al., 2016; Reyna et al., 2014). Experiment 1 shows that the relative evaluative polarity of the option in the decision problem, determined by the evaluative meaning of the words used to describe the options and their negated expressions, has a stronger impact on participants' behavior than the evaluative polarity of the problem, at least for the Versions using positive evaluative polarity words. In contrast to Fuzzy Trace Theory, we argue that it is the words that are used to describe the options that define the evaluative polarity of the problem, and that the evaluative polarity of the problem (positive or negative) influences the choice. In the ADP, the options occur in the context of a cover story profiled against the DISASTER frame. This frame is kept constant and is the point of departure from which all choices are made, and the backdrop against

which all the participants are making their decisions. What we vary is what is within the options presented to the participants. More precisely, we vary the way in which the options are presented, i.e. with and without negation, with numerals or vulgar fractions and the ordering of these (see Mandel, 2014 for a more precise approach to the numeric expressions, though). The fact that people actually interpret *saved* and *die* as non-scalar is crucial for the experiment since we are proposing a model based on an either-or understanding of the antonyms. In addition, when options were described with the positive descriptor *saved*, negated and non-negated, participants chose the risky option, as predicted if purely positive frames are preferred to mixed frames, i.e. that participants simply rely on the evaluative polarity of *saved* in order to assess the options. Contrary to our expectations, the same was not true for options described with the negative descriptor. As we have pointed out before, this may be due to the fact that there is a magnitude discrepancy in the descriptors' evaluative polarity. NEGATOR + *saved* is felt to be more negative than NEGATOR + *die* is positive. For instance, along the evaluative scale NEGATOR + *die* may be a good thing ('saved') or a very bad thing 'very badly injured', while NEGATOR + *saved* is likely to be straightforwardly understood as a bad thing ('die').

These findings and the linguistic approach to relative evaluative polarity allowed us to develop the account into the Extended-Relative-Evaluative-Polarity Account. In this, final account, the distance in evaluative polarity between negated and non-negated expressions used to describe options was used to predict the likelihood that participants will be risk averse or risk seeking. This was tested in Experiment 2 where we also varied the degree of evaluative polarity evoked by the two bounded, non-scalar, antonymic expressions, *saved* and *die*, by using versions of them. Experiment 2 demonstrated that differences in evaluative polarity also led to different degrees of risk preference. Participants do in fact respond to the evaluative polarity of the words used to describe the options. Therefore, it is not only the positive or negative gist of the expression as Fuzzy Trace Theory would have it, but also the evaluative polarity (how positive or negative that gist is) that affects risk seeking in the framing effect. We believe that the Extended-Evaluative-Polarity Account can act as an important complement to Fuzzy Trace Theory, and agree with Mandel that natural language interpretations of the framing effect must be taken seriously in the design and analysis of experimental data.

As we have noted, it is interesting that both Experiment 1 and 2 provided stronger effects for versions using mainly positive evaluative polarity words than for versions using mainly negative evaluative polarity words (see Figs. 5 and 6). This ties in with Kühberger's and Gradl's (2013) finding that framing only influences the safe option, i.e. Program A. They observed, as did Peters and Levin (2008), that when participants were asked to evaluate the attractiveness of the different programs, the Versions produced clear differences in attractiveness, so that Program A was considered more attractive in Version 1 than in Version 2, but, unexpectedly, this did not hold for Program B. At this point in time, we have no explanation for this. It is an issue that deserves to be looked into more closely in future studies.

Words have emotional effects and these may, to some extent, be measured numerically (e.g. Bradley and Lang, 1999). Our account makes use of a simple algebraic model over the space of word values. This algebraic model is similar to Anderson's (1981) information integration theory but with weaker assumptions. Indeed, we did find that the distance in evaluative polarity between positive and negative verbal expressions and their negations (NEGATOR + *die*) to some extent can be used to predict the probability of participants' risk seeking behavior (for negative differences) and risk averse behavior (for positive differences) respectively. This can be seen as a form of "loss aversion" (Kahneman and Tversky, 1979) in the space of word values but not in the space of outcomes. We doubt that the experimental method used in Experiment 2 can be further extended, however. As we have already stated, there is not much room for stretching the meanings of words related to death and life, since they are bounded, either-or, meanings, for which there is not much room for subjective interpretations along dimension of more and less, as would have been the case, had we used words such as *short* and *long* (Paradis, 2001; Paradis and Willners, 2006). There is yet another potentially more powerful problem involved, which concerns the very wording of the options of the cover story in ADP viewed in terms of experimental design. This problem has to do with consistency across the alternatives (see Kühberger and Gradl, 2013, for a related point).

In Version 1, the expressions are *saved* and *no people will be saved*, while in Version 2 they are *die* and *nobody will die*. The discrepancy across the two versions in terms of the expressions is of course not an ideal ground for an experimental design, nor is the fact that one of the word meanings presupposes some kind of agentive intrusion (*saved*), while the other does not (*die*). Now, the main goal of the present study was to replicate Tversky's and Kahneman's (1981) experiment in order to evaluate the participants' inclinations to select the risky or the safe option, assuming that it is *not* the reference point that is crucial for the participants' decisions but the evaluative polarity of the words. For this study to be a replication of the original ADP, we had to keep their formulations of the different versions of the cover story intact. However, in future research it would clearly be interesting to work out a linguistically controlled design of the formulation of the different version.

An interesting continuation of this research would be to see to what extent our observations generalize to other types of framing, such as attribute framing, in which case, evaluative polarity will not be related to how risk seeking participants are, but rather, for instance, to what extent they trust information given. Hilbig (2009) demonstrated that participants are more likely to judge a statement, such as "the percentage of successfully completed instances of attempted rapes", as being more trustworthy than those given the statement "the percentage of attempted but unsuccessful rapes". Could the extent

to which participants trust information instead be accounted for through the number of positively or negatively valenced words in each description, as in our Relative-Evaluative-Polarity Account?

A remaining problem in this line of research is why evaluative polarity should matter more to decision makers than the domain of the options. At least two possible explanations can be given, one concerning the structure of the decision task, and one concerning the role of emotions in decision making. Firstly, while it may be difficult to assess which domain an option belongs to, the triggering effect of the evaluative polarity of the words used to describe each option may be the solution to this problem. It is well known that decision makers are more strongly influenced by easily evaluated aspects than they are of less easily evaluated aspects (Hsee and Zhang, 2010) and by mechanisms related to the expectations in actual communicative situations. Secondly, the evaluative polarity of the words used to describe the options can also directly affect how pleasant each option is deemed to be. In real life, pleasantness, or lack thereof, may very well be excellent cues to whether we are dealing with potentially threatening situations, involving losses, or situations that are safe, involving gains. The strong link between decision making and emotional responses is well documented, which is very important since, otherwise, it would not be possible to evaluate the utility of the options in the choice situation. For instance, Simon (1983) states that if we set out to explain human rationality we have to take the role of emotions in choice behavior seriously. This link has been discussed in various fields such as in the risk-as-feelings literature (e.g. Loewenstein et al., 2001), dual process theories (e.g. Evans, 2008), or in research on how the expected hedonistic consequences of decisions are evaluated (e.g. Damasio, 1994). Our account fits well with such analyses.

Finally, it is clear that, while our account is still preliminary, our results are in agreement with other approaches that take both lexical meaning and the communicative intentions of the verbal descriptions that produce the framing effect into account. For example, McKenzie's (2003) review of how a Bayesian perspective that considers the relevance of frames can explain the framing effect – yet, not results such as those of the ADP – and Kühberger's (1995) ideas that probabilistic mental models of frames (Gigerenzer et al., 1991) can explain the ADP problem results. Fuzzy trace theory (Reyna and Brainerd, 1991; Chick et al., 2016; Reyna et al., 2014) is yet another approach that explains framing effects in terms of two types of representations of word meaning, called 'gist' and 'verbatim'. Kühberger and Tanner (2010) convincingly point to the need of combining this type of theoretical approach with notions such as information leakage in order to truly understand what is going on in tasks such as the ADP. We may conclude that Prospect theory, which is a formal theory to begin with, is simply not rich enough to explain what is going on. To this we add the importance of understanding the linguistic subtleties of the ADP task relative to the evaluative power of the words, the role of negation and combinations of descriptions in different constellations which in turn have different strengths of evaluative polarity in their wake.

By putting these pieces together, we hope to have made headway in our endeavors to develop a theory of the framing effect where the evaluative contributions of words play an important part. If it turns out to be the case that attractive, positive meanings have more power in decision making more generally than their negative counterparts, then new avenues open up regarding the importance of participants' emotional sensitivity to wording in decision making and the relative power of positive descriptors over negative ones.

Acknowledgements

We are very grateful to Florian Steinmann for his work on a previous version of this manuscript, to Joost van de Weijer for his generous help with statistical analysis, to Susanna Bernstrup and Alexander Sturkeltj for helping us collect data. We thank Fredrik Heinat and Philip Pärnamets for their comments on an earlier draft. We also thank Magnus Bergwallsfonden, Riksbankens Jubileumsfond and the Max Planck Institute for their financial support. In addition it is a pleasure to thank our anonymous reviewers that have provided insightful and helpful comments throughout the process.

References

- Anderson, Norman H., 1981. *Foundations of Information Integration Theory*. Academic Press, New York.
- Bradley, Margaret M., Lang, Peter J., 1999. *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings*. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.
- Chick, Christina F., Reyna, Valerie F., Corbin, Jonathan C., 2016. Framing effects are robust to linguistic disambiguation: a critical test of contemporary theory. *J. Exp. Psychol. Learn. Mem. Cogn.* 42 (2), 238–256.
- Damasio, Antonio R., 1994. *Descartes's Error: Emotion, Reason and the Human Brain*. Putnam, New York.
- Evans, St. John B.T., 2008. Dual-processing accounts of reasoning, judgment and social cognition. *Annu. Rev. Psychol.* 59, 255–278.
- Fagley, N.S., 1993. A note concerning reflection effects versus framing effects. *Psychol. Bull.* 113, 451–452.
- Fillmore, Charles, 2006. Frame semantics. In: Geeraerts, D. (Ed.), *Cognitive Linguistics: Basic Readings*. Mouton de Gruyter, Berlin and New York, pp. 373–400.
- Fischhoff, Baruch, 1983. Predicting frames. *J. Exp. Psychol. Learn. Mem. Cogn.* 9, 103–116.
- Fujii, Satoshi, Takemura, Kazuhisa, 2003. *Contingent Focus Model of Decision Framing Under Risk*. Technical Report 67, Tokyo Institute of Technology, pp. 51–67.

- Gigerenzer, Gerd, Hoffrage, Ulrich, Kleinbölting, Heinz, 1991. Probabilistic mental models: a Brunswikian theory of confidence. *Psychol. Rev.* 98, 506–528.
- Giora, Rachel, 2006. Anything negatives can do affirmatives can do just as well, except for some metaphors. *J. Pragmat.* 38 (7), 981–1014.
- Giora, Rachel, Balaban, Noga, Fein, Ofer, Alkabets, Inbar, 2005. Negation as positivity in disguise. In: Colston, H., Katz, A. (Eds.), *Figurative Language Comprehension*. Laurence Erlbaum, Mahwah, NJ, pp. 233–258.
- Giora, Rachel, Fein, Ofer, Aschkenazi, Keren, Alkabets-Zlozover, Inbar, 2007. Negation in context: a functional approach to suppression. *Discourse Process.* 43, 153–172.
- Giora, Rachel, Heruti, Vered, Metuki, Nili, Fein, Ofer, 2009. When we say no we mean no: on negation in vision and language. *J. Pragmat.* 41, 2222–2239.
- Giora, Rachel, Fein, Ofer, Metuki, Nili, Stern, Pnina, 2010. Negation as a metaphor-inducing operator. In: Horn, L. (Ed.), *The Expression of Negation*. Mouton de Gruyter, Berlin, pp. 225–256.
- Hilbig, Benjamin E., 2009. Sad, thus true: the negativity bias in judgments of truth. *J. Exp. Soc. Psychol.* 45, 983–986.
- Holleman, Bregje C., Maat, Pander, Henk, L.W., 2009. The pragmatics of profiling: framing effects in text interpretation and text production. *J. Pragmat.* 41 (11), 2204–2221.
- Horn, Laurence, 1989. *A Natural History of Negation*. Chicago University Press, Chicago.
- Horn, L. (Ed.), 2010. *The Expression of Negation*. Mouton de Gruyter, Berlin.
- Hsee, Christopher K., Zhang, Jiao, 2010. General evaluability theory. *Perspect. Psychol. Sci.* 5 (4), 343–355.
- Jones, Steven, Murphy, M. Lynne, Paradis, Carita, Willners, Caroline, 2012. *Antonyms in English. Construals, Constructions and Canonicity*. Cambridge University Press, Cambridge.
- Kahneman, Daniel, Tversky, Amos, 1979. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica* 47, 263–292.
- Kessler, Eric H., Ford, Cameron M., Bailey, James R., 1996. Object valence as a moderator on the framing effect on risk preference. *J. Econ. Behav. Organ.* 30, 241–256.
- Kühberger, Anton, 1995. The framing of decisions: a new look at old problems. *Organ. Behav. Hum. Decis. Process.* 62, 230–240.
- Kühberger, Anton, Grادل, Patricia, 2013. Choice, rating and ranking: framing effects with different response modes. *J. Behav. Decis. Making* 26, 109–117.
- Kühberger, Anton, Tanner, Carmen, 2010. Risky choice framing: task versions and a comparison of Prospect theory and Fuzzy trace theory. *J. Behav. Decis. Making* 23 (3), 314–329.
- Levin, Irvin P., Schneider, Sandra L., Gaeth, Gary J., 1998. All frames are not created equal: a typology and critical analysis of framing effects. *Organ. Behav. Hum. Decis. Process.* 76, 149–188.
- Loewenstein, George F., Hsee, Christopher K., Weber, Elke U., Welch, Ned, 2001. Risk as feelings. *Psychol. Bull.* 127 (2), 267–286.
- Mandel, David R., 2001. Gain-loss framing and choice: separating outcome formulations from descriptor formulations. *Organ. Behav. Hum. Decis. Process.* 85, 56–76.
- Mandel, David R., 2014. Do framing effects reveal irrational choice? *J. Exp. Psychol. Gen.* 143 (3), 1185–1198.
- Markowitz, Harry M., 1959. *Portfolio Theory: Efficient Diversification of Investments*. John Wiley and Sons, New Jersey.
- Maule, John A., 1989. Positive and negative decision frames: a verbal protocol analysis of the Asian Disease Problem of Tversky and Kahneman. In: Montgomery, H., Svenson, O. (Eds.), *Process and Structure in Human Decision Making*. Wiley, Chichester, UK.
- Maule, John A., Villejoubert, Gaelle, 2007. What lies beneath: reframing framing effects. *Thinking Reason.* 13, 25–44.
- McKenzie, Craig R.M., 2003. Rational models as theories—not standards—of behavior. *Trends Cogn. Sci.* 7, 403–406.
- McKenzie, Craig R.M., Nelson, Jonathan D., 2003. What a speaker's choice of frame reveals: reference points, frame selection, and framing effects. *Psychonom. Bull. Rev.* 10, 596–602.
- Oganian, Yulia, Korn, Christoph W., Heekeren, Hauke R., 2016. Language switching but not foreign language use per se – reduced the framing effect. *J. Exp. Psychol. Learn. Mem. Cogn.* 42, 140–148.
- Paradis, Carita, 2001. Adjectives and boundedness. *Cogn. Linguist.* 12, 47–64.
- Paradis, Carita, 2005. Ontologies and construals in lexical semantics. *Axiomathes* 15, 541–573.
- Paradis, Carita, 2015. Meanings of words: theory and application. In: Haß, U., Storjohann, P. (Eds.), *Wort Handbuch and Wortschatz. Handbücher Sprachwissen-HSW, Band 3*. Mouton de Gruyter, Berlin, pp. 274–294.
- Paradis, Carita, Willners, Caroline, 2006. Antonyms and negation: the boundedness hypothesis. *J. Pragmat.* 38 (7), 1051–1080.
- Paradis, Carita, Willners, Caroline, 2011. Antonymy: from convention to meaning-making. *Rev. Cogn. Linguist.* 9 (2), 367–391.
- Paradis, Carita, Willners, Caroline, 2013. Negation and approximation of antonymic meanings as configuration construals in SPACE. In: Paradis, C., Hudson, J., Magnusson, U. (Eds.), *The Construal of Spatial Meaning: Windows into Conceptual Space*. Oxford University Press, Oxford, pp. 287–311.
- Peters, Ellen, Levin, Irwin P., 2008. Dissecting the risky-choice framing effect: numeracy as an individual-difference factor in weighting risky and riskless options. *Judgment Decis. Making* 3 (6), 435–448.
- Reyna, Valerie F., Brainerd, Charles J., 1991. Fuzzy-trace theory and framing effects in choice: gist extraction, truncation, and conversion. *J. Behav. Decis. Making* 4, 249–262.
- Reyna, Valerie F., Chick, Christina F., Corbin, Jonathan C., Hsia, Andrew N., 2014. Developmental reversals in risky decision making: intelligent agents show larger decision biases than college students. *Psychol. Sci.* 25, 76–84.
- Sher, Shlomi, McKenzie, Craig R.M., 2006. Information leakage from logically equivalent frames. *Cognition* 101, 467–494.
- Sher, Shlomi, McKenzie, Craig R.M., 2008. Framing effects and rationality. In: Chater, N., Oaksford, M. (Eds.), *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*. Oxford University Press, Oxford, pp. 79–98.
- Simon, Herbert A., 1983. *Reason in Human Affairs*. Stanford University Press, Stanford, CA.
- Tottie, Gunnell, 1991. *Negation in English Speech and Writing: A Study in Variation*. Academic Press, San Diego.
- Tottie, Gunnell, Paradis, Carita, 1982. From function to structure. Some pragmatic determinants in impromptu speech. In: Enkvist, N.-E. (Ed.), *Impromptu Speech*, vol. 78. Publications of the Research Institute of Åbo Academy Foundation, pp. 307–317.
- Tversky, Amos, Kahneman, Daniel, 1981. The framing of decisions and the psychology of choice. *Science* 211, 453–458.

- Yamagishi, Kimihiko, 1996. [Strength of preference and effects of valence in the domains of gains and losses](#). *Organ. Behav. Hum. Decis. Process.* 66, 290–306.
- Yamagishi, Kimihiko, 2002. [Effects of framing and valence in decision-making: assessing decision-makers' perceived domains of choice](#). *Jap. Psychol. Res.* 44, 209–227.

Annika Wallin is associate professor in Cognitive Science, Lund University. In her work she focuses on decision-making in everyday life. She is interested in the relationship between normative and descriptive models of decision-making.

Carita Paradis is Professor of English Linguistics at the Centre for Languages and Literature at Lund University, Sweden. Her main research interests concern meaning in the broad sense. She specializes in lexical semantics and the modeling of meaning within the Cognitive Linguistics framework.

Konstantinos Katsikopoulos is Senior Researcher at the Center for Adaptive Behavior and Cognition of the Max Planck Institut für Bildungsforschung, Germany, and Associate Professor at the Department of Decision Analytics and Risk of the University of Southampton, UK. He does theoretical and applied work on integrating standard decision theory with the simple rules of thumb people use.