

Alive or not dead: Implications for framing from research on negations

Yaacov Schul

Hebrew University of Jerusalem

This is a close-to-final draft of the chapter in

G. Keren (Ed.) Perspectives on Framing. NY: Psychology Press (2011), pp 157-176

Preparation of this chapter was supported by grants from the Israel Science Foundation. I would like to thank Ruth Mayo for her comments on an earlier draft of this chapter.

Manipulations of framing might be conceived as a means of inducing two ways of viewing an issue – as if the issue is approached from two different directions. From this perspective the existence of framing effects might be interesting because of what is known about the effect of perspective on perception. To wit, although the retinal image of an object seen from different locations is different, the mind typically normalizes the image, so that perception is relatively invariant to the exact angle of view. Why, then, does thinking about the same issue from different conceptual angles – saving vs. losing lives, including vs. rejecting options – leads to different decisions? Why is the evaluation of decision problems different from the perceptual categorization of physical objects?

One difference between perception and decision has to do with people's assumptions about the two kinds of phenomena. Physical objects are assumed to be perception-invariant entities. People are well aware that they may see different aspects of an object when they view it from different locations. They attribute these differences to the conditions of perception, assuming that the objects themselves are stable. Hence, the mental system has developed normalization procedures to correct for variations attributed to the perceptual conditions.

However, contrary to the normative requirement of decision invariance (Tversky & Kahneman, 1981), decision issues are not invariant to conceptual angles because people are aware that their meaning often includes hidden attributes. As a result, their meaning has to be negotiated between the parties who espouse the different angles. To illustrate, consider two decision makers, a buyer and a seller, who are looking at the same car. It is likely that they interpret the same percept quite differently. For instance, for a seller, a shiny red car may mean a car in a mint condition. For a buyer, the same car may imply "look under the hood, something must be wrong". Indeed, more than half a century ago, Heider and Simmel (1944) demonstrated that a triangle following a circle on a filmstrip might be interpreted either as a circle leading a triangle or as a circle escaping from a triangle. More recently, Blakemore and Decey (2001) suggested the existence of brain structures

which attempt to uncover intentions when human actions are perceived. Interestingly, these structures are not triggered when the actions are attributed to non-humans. These examples suggest that social perceivers are well aware that in attempting to decipher the physical events and assign them meanings, they must use, in addition to the characteristics of the events themselves, subtle cues produced by the environment. Framing effects may result from the utilization of such external cues when people are trying to make sense of the event.

Specifically, according to this perspective, framing effects stem from the existence of two sets of extra-stimulus cues, which are differentially triggered under the two frames. For example, thinking about exclusion means something slightly different than thinking about inclusion (Yaniv & Schul, 1997, 2000). Although the dictionary definition of “exclusion” and “inclusion” might be complementary, so that “exclude” might be defined as “not include”, the social connotations of these terms might be different. People know that in everyday language the two terms are not interchangeable, and that there are circumstances under which communicators choose to describe an action as inclusion rather than exclusion. Similarly, although the dictionary definition of saving and losing lives might be complementary (in the sense that not-alive means dead and not-dead means alive), the network of associations which accompany “life” and “death” makes the two concepts different. Thus, it is often the case that people who are probed directly about the meaning of two complementary words or phrases, which serve to frame an issue, argue that the two are exact opposites, thus being logically equivalent. Yet, each of the frames contains information that allows decision makers to translate the logically-equivalent problems into psychologically-nonequivalent ones (cf., Sher & McKanzie, 2006).

There is, however, an additional factor that contributes to differences in decisions made under alternatives frames. Two frames do not consist merely of two opposite perspectives, such as life saved vs. life lost, or include vs. exclude. Under each frame, decision makers must consider whether to act in a particular way (e.g., save lives, include an option) or not. To illustrate, consider

the inclusion/exclusion framing manipulation (e.g., Yaniv et al, 2000, Levin et al, 1998). People might be asked which options to include in one framing condition, and which ones to exclude in another framing condition. Although decision makers are being asked either to include or to exclude options, in all likelihood they are doing something slightly different. In the inclusion condition they consider which options to include and which ones not to include. Similarly, in the exclusion condition they ask themselves which options to exclude and which not to. Why is this reformulation important? Because the meaning of the negations (e.g., the meaning of “not include” and “not exclude”) may contribute to the generation of the framing effect. This is described schematically in Figure 1.

Panel (a) of Figure 1 portrays what might go on in one's mind as one considers which options to include. For example, assume that the decision maker is a groom who has to decide which family members to invite to his wedding. To do so, the groom imagines the type of person who is definitely worth inviting and the type of person who should definitely not be invited. Somewhere in the middle, between these two types, the groom forms a demarcation line – a criterion for invitation. Whoever is to the left of the criterion on this imaginary mental scale is included in the guest list.

Insert Figure 1 About here

The other panels describe the groom's father's way of thinking. The father, in contrast to the son, considers whom to exclude from the guest list. Again, he forms a mental scale ranging from “exclude” to “not exclude”, and sets up a criterion for exclusion. Whoever is to the right of the criterion is excluded, and the rest are not excluded, and therefore implicitly included. In this formulation, the interpretation of not X becomes important. By changing how not X is interpreted one can vary the regions of inclusion and not-exclusion and model various forms of framing effects.

Panel (b) depicts the case when the positioning of the “not-include” criterion is identical to that of the “not-exclude” criterion. As a consequence, there is no framing effect – the set of guests

selected by the groom coincides with the set selected by the father. Panel (c) depicts the traditional framing effect (e.g., Yaniv & Schul, 1997, 2000), whereby more guests are invited by the father (the non-exclusion region) than by the son (the inclusion region). Finally, panel (d) portrays the opposite outcome, in which the non-exclusion region is smaller than the inclusion region.

These examples show that the magnitude of framing effects can reflect how decision makers interpret negations. Negation might be interpreted as the opposite term ("not-include" means "exclude"), a mitigated version of the opposite term ("not-include" is not as extreme as "exclude"), or as a stronger version of it. In the rest of this chapter I describe several studies, all of which were designed to investigate how people process and interpret messages phrased as negations. I shall focus on two types of contrasts: either comparing the nature of associative processes triggered by affirmations (e.g., "Adam is lazy") and negations ("Adam is not lazy"), or comparing how the negation of a term is understood relative to the opposite term. Specifically, assume that X and opposite X are two complementary terms, such as "good" and "bad". Our research investigated the extent to which the negation of one alternative (e.g., not-good) resembles the opposite alternative (bad) in meaning, and, more importantly, what determines the magnitude of discrepancy in meaning between the two.

The nature of the associations activated by negations

What do people think when they hear a negated communication such as "Ben is not a romantic person"? Mayo, Schul, and Burnstein (2004) distinguished two kinds of attributes, which they termed bipolar and unipolar. A bipolar attribute has a well-defined opposite construct which is easily accessible (e.g., *hot* brings to mind *cold*), whereas a unipolar attribute does not (e.g., there is no readily available opposite term for being *romantic*). For ease of presentation we refer to the bipolar/unipolar distinction as if it forms a dichotomy. It is clear, however, that variations in the readiness with which opposite associations come to mind give rise to a continuum. In order to study the effects of this readiness, Mayo et al. selected good exemplars from the extreme ends,

either adjectives that had highly available and well-defined opposites (bipolar) or those which fail to trigger a well-defined opposite term (unipolar). I return to the question of the prevalence of such extreme cases after describing our main findings.

What are the implications of the bipolar/unipolar classification? Since one can readily access an opposite mental construct when processing the negation of a bipolar description, one is likely to spontaneously activate associations that are congruent with the opposite. Accordingly, the associations triggered by a negation are congruent with the intended meaning of the negation. To illustrate, upon hearing that "Adam is *not smart*", one might spontaneously think of stupidity-related associations such as dumb acts or stupid persons. Unipolar descriptions, by definition, do not have a readily available opposite schema. Their negations, therefore, tend to trigger associations and implications of what has been negated ("*Adam is not romantic*" may trigger behaviors that are romantic, which are then denied or negated). Accordingly, these associations are incongruent with the intended meaning of the negation.

To examine this prediction, Mayo et al. showed participants, on each trial, a description of a person which was either affirmative (e.g., "Adam is romantic") or negative (e.g., "Adam is not romantic"), and either bipolar or unipolar. Subsequently, participants read a statement about a behavior that was either congruent with the trait (e.g., "wrote a love letter"), or congruent with the negation of the trait (e.g., "failed to remember a birthday"), or an unrelated behavior. Participants had to indicate whether the behavior was congruent with, incongruent with, or unrelated to the person description. Overall, participants made very few mistakes, suggesting that the degree of congruence of the behaviors was unambiguous.

Our primary analysis involved the speed of classifying the behavior as congruent or incongruent. We assumed that participants' classification would be faster if the behavior corresponded to the kind of associations triggered spontaneously by the person descriptions. Thus, we predicted, and found, that the tendency of a negation to spontaneously activate incongruent

associations would depend on the type of information being negated. Incongruent behaviors were facilitated by negations of unipolar adjectives, whereas congruent behaviors were facilitated by negations of bipolar adjectives. This suggests that whereas unipolar negations (e.g., "not romantic") are represented as notX, bipolar negations are represented as oppositeX.¹ Put differently, bipolar negations tended to be translated into the opposite term, while unipolar negations were not.

The above conjecture is also supported by an analysis of the participants' memory for the descriptions. Whereas memory for affirmations involving bipolar traits was similar to memory for affirmations involving unipolar traits, memory for negations involving the two types of traits involved different types of mistakes. Specifically, mistakes of dissociation (e.g. remembering "not responsible" as "responsible") were more prevalent in the memory reports of the unipolar traits than in the reports of the bipolar traits.

Taken together, the pattern of judgment latency and recall mistakes is consistent with the suggestion that the phrasing of a description (affirmative versus negative) and the nature of the description (with or without a readily available opposite schema) determine the impact of a negated message. Accordingly, Mayo et al. (2004) proposed that the existence of a schema that accommodates the meaning of the original negation is critical in determining how a negation will be encoded.

The prevalence of unipolar negations

In constructing the stimulus material for the experiments reported in Mayo et al. (2004), it became clear that, more often than not, adjectives are bipolar. Does it mean that unipolar descriptions are esoteric phenomena, devoid of empirical significance? I propose that the answer is "no". When descriptors are behavioral or episodic, they often act as if they are unipolar. "He did not have sex with that woman" cannot be translated readily into an affirmation – there is no clear

¹ When the descriptions, be they bipolar or unipolar, were phrased affirmatively (e.g., "Adam is romantic"), the congruent behaviors were classified more quickly than the incongruent behaviors.

schema that accommodates what he did have with the woman because he could have done many things, each of which might be described as "not having sex". Similarly, the statement "there is no bird in the sky" (Kaup, et al., 2006) is unipolar since we know nothing about what is in the sky. More generally, unipolar mental objects might be contrasted either with many specific alternatives none of which dominates the others or with something vague that is best defined as not-being—the unipolar concept. In such cases people are likely to activate associations which are in line with the negated concept, and therefore, incongruent with the intended meaning of the negation. As a result, judgments based on X and not X are unlikely to be complementary.

Implications for framing

The presumed variation in the associative structure triggered by negations and affirmations made little difference for the congruency judgments themselves, in the sense that the accuracy of the judgments was uniformly high whether the description was affirmative or negative and whether it included bipolar or unipolar adjectives. This may not be surprising given the very crude nature of the behavioral probes: behaviors were selected to be either highly congruent, highly incongruent or completely unrelated. Such a classification task does not allow for tests of the shades of meaning within a category. Yet, the latency data suggest that people were activating different kinds of associations in processing different types of negations. In particular, Mayo et al. proposed that negations of unipolar traits are best described by a schema-plus-negation-tag model. For example, upon hearing that someone is "not romantic" people may think about romance-related behaviors (e.g., "bringing her flowers") and add a tag denoting the rarity or absence of the event ("he never brings her flowers"). Because X-related associations were triggered, not X becomes a weakened or mitigated version of opposite X. Below I discuss the mitigation of negations in greater detail.

Let me mention briefly that the weakened persuasive potential of messages phrased as notX might be particularly important in persuasion campaigns. Consider a defense attorney who is trying to exonerate a client accused of a murder. The message, "he is not a murderer" would be less

persuasive than the complementary claim by the prosecution "he is a murderer", because, as I suggested above, the seeds of believing in X are implemented by arguing that someone is notX, especially when X is unipolar. To cope with such a handicap, the defense must provide an alternative schema for interpreting the testimonies and summarizing the evidence. One useful strategy involves the introduction of the "He was framed" schema. Such an alternative schema allows the decision maker to reinterpret the evidence in a new light, as showing not only that the accused is not a murderer, but also that the evidence can be interpreted in line with innocence. This should be more effective than merely denying the claim.

The meaning of negated bipolar adjectives

The discussion of Mayo et al.'s research highlighted the contrast between unipolar and bipolar adjectives. The research of Fraenkel and Schul (2008) used bipolar adjectives to investigate the kind of inferences people make from negations with a more refine measure of meaning. Briefly, this research suggests that even if one restricts the analysis to bipolar adjectives, the meaning of the negation of an attribute X (notX) is weakened compared to the meaning of the opposite of X (oppositeX). Our theoretical analysis attributes this effect to people's knowledge about the circumstances that lead communicators to use negations.

Let us start by asking why people use negations in communication even in cases the message can be conveyed by using the corresponding opposite term affirmatively? This is puzzling because processing negations is often harder, slower, and more error-prone than processing affirmations (e.g., Carpenter & Just, 1975; Just & Clark, 1973; Lea & Mulligan, 2002; MacDonald & Just, 1989; Wason, 1965). Past research has suggested several reasons, which are not mutually exclusive. Importantly, these reasons have to do with the meaning of negation in interpersonal interactions. Negations might be used as a means of being polite (Colston, 1999; Giora, et al., 2005; Horn, 1989). That is to say, social norms of communications sometimes dictate what can and cannot be said. Hence, in certain contexts, communicators can avoid being perceived as blatant by

negating an adjective instead of using an affirmation which has a strong derogatory connotation (e.g., saying "John is not smart" is more polite than saying "John is stupid"). Negations might also be employed when one wants to contradict a common expectation or belief held by the receiver (Clark & Clark, 1977; Givon, 1978; Jordan, 1988; Wason, 1965). Fraenkel and Schul (2008) tested a third, related alternative. They hypothesized that when people want to convey a mitigated sense of an adjective, they negate its opposite.

Fraenkel and Schul proposed that the mitigated sense of negations can be driven by two fundamentally different mechanisms that are involved in (i) encoding a negated adjective and (ii) interpreting it as an act in an interpersonal interaction. Schematically, encoding a negated statement could be described as proceeding in two phases. During the initial phase of comprehension, recipients activate the associations of each linguistic element in the message. Accordingly, when encoding the statement "The coffee is not hot", recipients activate inferences that are associated with "coffee", "hot", and potentially also associations which are triggered by the "not" operator (e.g., Alia-Klien et al., 2007; Hasson & Glucksberg, 2006; Giora, 2006; Kaup et al., 2006). As they continue processing, however, recipients integrate the different words in the statement and in doing so they can activate inferences that are congruent with "not hot" or "cold" (Fischler, Bloom, Childers, Roucos & Perry, 1983; Gannon & Ostrom, 1996; Mayo, et al, 2004; Whitmeyer, 1997).² The simultaneous presence of associations to "hot" and "cold" may induce a moderated meaning of the negated concept, with "not hot" being understood as neither very hot nor very cold.

The mitigation of the meaning of a negated message may also stem from the pragmatic inferences recipients make during the interaction. According to Grice (1975) and Horn (1984), communicators prefer expressions that incorporate all of the information they want to convey and at the same time are parsimonious. As a result, upon hearing a non-parsimonious communicative element such as negation, the listener often assumes that it has been generated for a purpose. Such

²Note that the findings of Mayo et al. (2004) suggest that in the case of unipolar negations people might be less likely or even completely unable to access inferences that are congruent with the intended meaning of the negation

conversational implicatures allow one to go beyond the literal meaning of the statement and provide a rich field of shades of meanings (Horn, 2004). For example, upon hearing the statement "The coffee is not hot" the listener may wonder why the affirmative statement, which has a simpler form, was not used. The listener may conclude that negation was used purposely to convey a weakened sense. Accordingly, "not hot" is understood to mean neither hot nor cold, but something in between. Since the two mechanisms operate at different stages of processing, they are not inconsistent with each other. In fact, I believe that both work in tandem, giving rise to the mitigated meaning accorded to negated statements.

Moderation of the mitigation of meaning

Assume the existence of two adjectives, X and oppositeX, with opposite meanings. When they span a range of meanings (e.g., from *warm* to *cool*, from *alive* to *dead*), we can ask about the sub-range of meanings of the pair X and notX. Meaning mitigation implies notX is less extreme than oppositeX, and therefore, that the difference between X and notX (e.g., *warm* and *not-warm*) is less than that between X and oppositeX (*warm* and *cool*). Mitigation is shown in Figure 1 by the distance between notX and oppositeX (e.g., *not-include* and *exclude*).

Fraenkel and Schul (2008) considered two factors as potential moderators of the extent of meaning mitigation. The first is a property of the adjective pair. A pair of antonymic adjectives are considered contraries if they can be used in two sentences that cannot be simultaneously true but may be simultaneously false (e.g., the coffee is cold or hot, or neither hot nor cold). In other words, contraries span a non-empty range of intermediate values. To illustrate, coffee might be hot, warm, lukewarm, cool, or cold. Contradicories, in contrast, are antonyms which follow the *law of excluded middle*, meaning that they can be used in two sentences, exactly one of which must be true (e.g., the door is open or closed, but the door must be either open or closed). This distinction is quite old, dating back to Aristotle, but also appearing in more recent essays (for example, Ladusaw, 1996).

If a specific object or event can be viewed *strictly* as either X or oppositeX, then, logically, if it is not-X, it must be oppositeX, with little room for mitigation. From this perspective, the mitigation hypothesis does not make sense with truly contradictory pairs. However, a little reflection suggests that contradictory pairs in everyday language are not quite that dichotomous. Just anyone can come up with possible realistic interpretations in which an apparent logical dichotomy has meaningful mid-values. Terms like "half dead" or "barely alive", "almost right" or "not quite wrong" are easily understood and often used in conversations. Thus, Fraenkel and Schul (2008) pondered whether differences between contraries (i.e., pairs of adjectives that span a range of values) and contradictions (i.e., dichotomous adjectives) would be revealed in people's judgments.

The second moderator of the meaning mitigation of negations has to do with the markedness of the adjectives. The *unmarked* member in a pair of adjectives is the usual, the normal, the common, and the neutral or less specific, compared to the *marked* member. Conversely, the *marked* member is the adjective which carries a distinctive feature that distinguishes it from the other member (Battistella, 1996; Boucher & Osgood, 1969; Hartmann & Stork, 1972; Holleman & Pander Maat, in press ; Levinson, 1983). The term markedness has been found useful not only as a means of understanding linguistic structures, but also for explaining errors in second-language learning (Santos, 1987), or even the organization of kinship universals (Hage, 2001).

How would markedness influence the mitigation of negation? Fraenkel and Schul (2008) proposed that the negation of an unmarked adjective should be closer in meaning to the opposite adjective than negation of a marked adjective. To illustrate, consider the good/bad adjective pair, and assume that "good" is unmarked and "bad" is marked (Rozin, Berman, & Royzman, in press). The above hypothesis means that "not good" resembles "bad" more than "not bad" resembles "good". Translating this into mitigation terminology, the hypothesis implies that the meaning of

“not good” is mitigated to a lesser extent than the meaning of “not bad”. Later on I shall discuss markedness in more detail, and in particular I shall elaborate on whether or not markedness is context-dependent.

There are several factors that can drive the proposed moderation by markedness. They hinge on the intuition that the unmarked member in an adjective pair provides the more common term for referring to a phenomenon. The marked member is therefore indicated by the absence of the normal term, by being not-unmarked. From this perspective, the use of the marked adjective indicates something special, since under ordinary conditions communicators tend to use the unmarked or normal term. Consequently, the absence of a marked quality does not eliminate the choice of the non-conventional term, so that negation of a marked term does not make it similar to the unmarked term. For example, assume that *tall* is the unmarked member of the tall/short pair. Assume further that you hear that someone is described as *not short*. You ponder why she was described in terms of shortness rather than tallness. One possible inference is that although she is not really short, she must be somewhat short, otherwise, why use the term "short" in describing her.

Fraenkel and Schul (2008) discussed a related mechanism based on considering the range of meanings associated with the unmarked and the marked adjectives. Since the marked adjective is mainly the opposite of its [unmarked] antonym, it should more closely resemble the negation of the unmarked adjective. The unmarked adjective has a broader meaning (being not merely the negation of the marked antonym), and therefore it should resemble the negated marked antonym less closely.

The analysis so far assumed that markedness is a property of adjectives. However, it is likely that markedness is also sensitive to the context, since the unmarked term is, by definition, the normal, common, or expected attribute in a particular setting (Holleman & Pander Maat, in press). It is therefore quite possible that what is normal in one setting can be abnormal in another setting. Consider, for instance the hot/cold pair. Commonly one asks how *hot* the soup is, but not how *hot* the ice cream is. Hotness might be unmarked in the context of soups (and summer days), but

marked in the context of ice creams (and winter days). Analogously, *happy* might be the unmarked member of the happy/sad pair in normal situations where happiness prevails (Matlin & Stang, 1978), but *sad* may become the unmarked adjective in situations where sadness prevails. Still, research shows that within a given culture people have good intuitions about cross-situational markedness. It is found that at least in Western societies, markedness is associated with evaluative negativity (Boucher & Osgood, 1969; Horn, 1989; Lehrer, 1985), so that negative qualities are seen as less common. I shall return to the significance of a context for markedness toward the end of this chapter.

Although markedness has been defined as a linguistic phenomenon, the above characterization of markedness makes it interesting to generalize it to non-linguistic domains by investigating its relationship with norms and exceptions (Kahneman & Miller, 1986). Speculatively, let me assume that normal events are analogous to unmarked concepts, while exceptions are analogous to marked ones. From this perspective it follows that negating a normal event makes it abnormal to a greater extent than negating an abnormal (or exceptional) event makes it normal. In part, such an asymmetry may have to do with the tendency to draw inferences from specific abnormal characteristics to abnormality in other aspects. This may remind the reader of the research on illusory correlations (see review in Fiedler, 2000). Accordingly, when people try to negate a specific characteristic, they fail to undo all the inferences that are associated with that characteristic (Schul & Burnstein, 1985; Schul & Mayo, 1999). If so, the asymmetry implied by markedness (vs. unmarkedness) reflects the difficulty people have in discounting correlated attributes, implying that mitigation should disappear under the conditions specified for successful discounting (Schul & Burnstein, 1998).

Fraenkel and Schul (2008), explored three questions: (i) Is there an overall mitigation of meaning of messages expressed as negations? (ii) Is mitigation more pronounced for contraries

(adjectives that lie on a continuum) than for contradictories (adjectives that form a dichotomy)? (iii) Is meaning mitigation of marked adjectives greater than that of unmarked adjectives?

The first experiment assessed people's direct judgments of the meaning of negations. Our respondents evaluated the resemblance of the meaning of a negated adjective and its antonymic counterpart. Pairs of sentences were presented on a computer screen, one above the other. One sentence contained a negated adjective (e.g., the coffee is *not hot*) and the other contained the adjective's affirmative antonym (e.g., the coffee is *cold*).³ Participants rated the resemblance in meaning of the sentences on a 21-point scale. We intentionally provided a minimal context in which each adjective or negation was embedded. This was done to balance the need to provide a pragmatic context for interpreting each adjective (or its negation) and the need to avoid constraining the meaning of the adjectives (or negations) by an elaborate context.

Table 1 presents the meaning resemblance of negations (*negX*) and the opposite terms (*oppositeX*). The entries are expressed as percentages, so that the higher the numbers the more similar a negation of X was to *oppositeX*. The table shows that, on average, the meanings of negations were mitigated as compared to the opposite terms. However, the overall extent of the mitigation was qualified by two interactions. There was a difference between contraries (adjectives that span a continuum, such as pretty-ugly, friendly-solidary, rich-poor) and contradictories (adjectives that form a dichotomy, such as alive-dead, right-wrong, real-fabricated). Whereas a negation of a contrary adjective (e.g., *not pretty*) was rated as being about halfway between the two contrary adjectives, a negation of a contradictory adjective was rated as quite similar to its antonym. Still, even in the latter case, *not X* was seen as a weakened version of *oppositeX*.

³ Half of the participants received the unmarked adjectives with the negation of the marked adjectives embedded in one pair of sentences, while the other half saw them embedded in the other pair of sentences (e.g., "The coffee was *hot*" "The coffee was *not cold*" and "The soup was *cold*" "The soup was *not hot*"). In addition, we created 12 filler pairs, each of which contained an adjective and the negation of an unrelated adjective. The fillers were used to anchor the lower end of the resemblance scale.

Table 1: Meaning resemblance as a function of type of adjective pair and markedness of the adjective (Fraenkel & Schul, 2008, Experiment 1)

	Contraries	Contradictries
Marked is negated	53.6%	81.0%
Unmarked is negated	48.8%	80.7%

Note: Higher numbers indicate that the negated adjective and the corresponding antonym were seen as more similar to each other, implying less mitigation of meaning.

The extent of mitigation was also affected by the markedness status of the adjectives when the two adjectives belonged to a contrary pair. A negation of the unmarked member of the pair was rated as less similar to its antonym ("not pretty" to "ugly") than a negation of the marked member ("not ugly" to "pretty"). This difference did not occur for contradictory adjectives.

A second experiment investigated the mitigation hypothesis with a different measurement procedure. We were inspired by a debate in the questionnaire-survey literature about the use of bipolar and unipolar scales (Gannon & Ostrom, 1996; Yorke, 2001). Briefly, bipolar scales have the scale endpoints labeled with two antonymic adjectives. In contrast, unipolar scales have the endpoints labeled with an adjective and its negation. Based on the mitigation hypothesis, we reasoned that the range of values covered by a unipolar scale should be smaller than that covered by a bipolar scale. Moreover, in line with our hypotheses about the adjective type and markedness, a mapping from a unipolar scale to the corresponding bipolar scale should depend on the type and markedness of the adjectives being used as the scale's endpoints.

Respondents in Experiment 2 were asked to mark the range covered by the anchors of a unipolar scale (for example, *hot – not hot*) on the corresponding bipolar scale (*hot – cold*). We manipulated the type and markedness of the adjective used for labeling the unipolar scale. Each participant was given a 60-page booklet. On each page the participant saw a bipolar scale (e.g., a scale anchored by "rich" and "poor") and the labels anchoring a unipolar scale underneath. These

labels consisted of one of the adjectives from the bipolar scale and its negated version (e.g., "rich" and "not rich"). Participants were asked to color (with a marker) a range on the bipolar scale which corresponded to the unipolar scale. Participants who thought that the negated adjective (e.g., "not rich") was very similar to the corresponding antonym (e.g., "poor") marked a larger portion of the bipolar scale than participants who thought that the negated adjective was a weakened version of the antonym. The findings were virtually identical with those found in the first Experiment.

How goals influence the processing of negations

Earlier it was proposed that in thinking about negations people apply their knowledge about communication in the social world. It follows that one can influence how negations are processed and communicated at a later time by manipulating the goals of communicators and/or how listeners construe the goals of the communicators. Fraenkel (2006) addressed this issue by examining whether the tendency of people to remember negations in their original form (i.e., recall notX as notX) rather than affirmatively as the corresponding antonym (recall notX as oppositeX) varies as a function of the social constraints of the interaction.

In one experiment participants heard a series of short stories, each of which contained a negation of an adjective that was either a member of a contrary pair or a member of a contradictory pair. After a short filler task, participants were given a typed version of the stories (each story appearing on a separate page). The stories included a few blank slots, and participants were asked to fill in the missing words – a cued recall task. Importantly, participants processed the stories under one of two roles. In one condition, participants were forewarned that they would be asked later questions about the stories and that they must be as accurate as possible in answering these questions. To emphasize this, participants were promised a monetary bonus based on the accuracy of their answers. Participants in a second condition were also warned that they would be asked questions about the stories. However, they were informed that they should strive for clarity so that

another person could understand the stories as accurately as possible. To emphasize this they were informed that the monetary bonus would be based on the clarity of their answers.

Fraenkel (2006) found that the prevalence of recalling a negation (notX) as the corresponding antonym (i.e., correctly translating the negation) varied a great deal between the two conditions. When the participants' goal was accuracy, negations were translated to affirmations in about 8% of the cases, with little difference between negations involving contrary and contradictory adjectives. However, when the participants were instructed to be clear, the prevalence of translation was more than double. Moreover, meaning-preserving translation was significantly more likely for negations of contrary terms. Thus, translation is influenced by meaning resemblance as well as by the constraints imposed by the social interaction.

More on the effect of context

The research described above, explored the cognitive outcomes of processing negations (notX). In particular, several different lines of evidence were discussed which suggest that people understand notX as implying a less extreme quality than that denoted by oppositeX. Our research revealed that the type of adjective (unipolar vs. bipolar), the type of adjective pair (contraries vs. contradictories) and the goals of the listener (accuracy vs. clarity) have robust effects on the interpretation of negation. The markedness of the adjectives that are being negated seemed to have a rather modest (albeit statistically significant) effect.

Markedness reflect an asymmetry in the use of the two alternatives, X and oppositeX. The limiting case of such an asymmetry involves unipolar adjectives, because the oppositeX alternative tends to be absent in normal interactions and thus can be specified only by negating the existing alternative. However, precisely because of this characteristic, unipolar adjectives are unlikely to be used in framing manipulations. Still, this limiting case points to the potential significance of asymmetry to the understanding of framing. The phenomenon of markedness refers to a less extreme asymmetry, namely, the case in which one of the alternatives is more prevalent and

therefore is considered normal (at least in the particular setting implied by the message), making the other alternative seems like an exception. We expected to find robust markedness effects on the interpretation of negations. However, markedness had only minor influence on the way negations were interpreted.

One may suspect that the size of the markedness effect obtained by Fraenkel and Schul (2008) is an underestimate of the influence of asymmetry in framing paradigms for two reasons. The first reason is somewhat technical and has to do with the measurement procedure. Fraenkel and Schul's (2008) study involved a within-participant comparison – namely, participants were presented with two opposing alternatives and mapped the negation to a point within the range which was anchored by these alternatives. As the two alternatives were highly salient, and the processing of negation was indicated by an explicit judgment of resemblance (of the negation to the opposite term) the findings might have been influenced by the participants' theories about the effect. This may not have been the case had the effect been tested in a between-participants design, as is typically employed in decision-framing studies.

The second possible reason for underestimating the effect of markedness in our studies may involve our attempt to minimize the influence of context. The processing of negations is influenced not only by the meaning of what is being negated, but also by the context in which the negation is expressed. Recipients of communications, listeners and readers alike, are not passive processors of information, and they pay a good deal of attention to what is being said as well as to how it is expressed. This practice is useful because one can learn quite a bit from the expressions being used, provided that one is familiar with what is normal in the particular setting. This, in turn, implies that the impact of extra-content cues, such as markedness, can vary between situations because the same cue may have a different meaning in a different context.

Specifically, because context affects what is normal or expected, it can influence which member of an adjective pair is the unmarked member and, as a result, how negations are understood. Consider, for example, the following two situational contexts:

- 1: Kim told her father that she had wrecked his car. Her boyfriend asked her how he took it and she said: (a la Colston, 1999)
- 2: Kim told her father that she had won a scholarship for college. Her boyfriend asked her how he took it and she said:
 a: He was mad
 b: He was not mad
 c: He was happy
 d: He was not happy

As Fraenkel and Schul (2008) noted, "Mad" (in a) is expected or normal in the context of (1); "happy" (in c) is normal in the context of (2). As the impact of negation depends on the normality or expectedness of the negated concept, the context should affect how "not mad" (in b) is interpreted. Specifically, in context (1), where "mad" is normal, it should resemble "happy" more than in context 2, where it is abnormal. This reasoning complements Colston's (1999) theoretical analyses and results, at least where negative adjectives are concerned. Context might influence framing by affecting the asymmetry between the two options, X and oppositeX. By communicating messages without a context we forced the participants in our studies to make an inference about a context, which might have introduced great variability between participants. To counter this problem the experiment could utilize a context that reinforces the asymmetry between the marked and unmarked adjectives. Holleman and Pander Maat (2009) have recently developed a theoretical framework for discussing how context influences markedness and thus framing, proposing that markedness depends on the direction of the goal implied by the context. To use a simplistic example, consider the "dead/alive" pair. If one is engaged in saving lives, "alive" becomes the

unmarked adjective, but when one is trying to exterminate lives (as in the case of roach extermination) then dead becomes the unmarked adjective. Participants in Holleman and Pander Maat study were given a scenario which set up an expectation (e.g., Next week I'll have little (vs much) time to come over and help you with odd jobs about the house). They were then asked to choose which of two alternative conclusions best fits the scenario: (i) My diary is half full, and (ii) My diary is half empty. Holleman and Pander Maat reported that participants chose more often the half-full conclusion as a continuation to the little-time than to the lot-of-time scenario, suggesting that although half-full and half-empty might be considered equivalent, respondents are sensitive to subtle cues in the context which lead them to distinguish between them. Thus, it might not be sufficient to consider the expectedness of the linguistic terms by themselves (i.e., compare unmarked and marked traits) but rather, linguistic terms should be studied within their function in the particular [social] context.

According to this theoretical analysis, the influence of context on markedness capitalizes on how people interpret events and actions -- in particular, on the type of outcome implied by a specific context. There is, however, a different mechanism through which context can influence the interpretation of adjectives in general, and negations in particular. The context may impose constraints on the kind of the language communicators can use, and listeners may take these constraints into account when they interpret the communication.

I have already referred to the norm of politeness which can make the use of a negation more or less appropriate and thus influence its interpretation. Let me discuss its effect in more detail. Imagine hearing a communicator describing someone, either favorably or unfavorably. The favorable description can be expressed affirmatively (e.g., "she is good") or as a negation (e.g., "she is not bad"). Analogously, the unfavorable description can be expressed affirmatively or as a negation ("she is bad" vs. "she is not good"). Due to politeness (Colston, 1999) and face-saving (Bonnefon & Willjoubert, 2006), listeners may reason that a communicator is using negation of a

favorable adjective in describing another person because asserting directly that a particular person has an unfavorable quality is not pleasant or polite and might be construed as face-threatening (rather than face-saving). From this perspective, therefore, negation of a favorable quality X may either mean a strong assertion about opposite X, or a mitigated version of it, expressed as notX. Such reasoning cannot be applied, however, to negations of unfavorable qualities. In this case, communicators can choose between describing the other favorably by using the opposite of the unfavorable quality (which is both polite and pleasant) or describing that person as not having the unfavorable attribute, that is, non-negatively. Since both acts of communication are socially appropriate, the choice between the two, and in particular, the use of negation, is informative. Because negations of unfavorable attributes are unambiguous, at least as compared to negations of favorable attributes, their meaning should be weakened to a greater extent. Thus, predictions from politeness and face-saving coincide with those based on markedness, assuming that marked adjectives tend to be evaluatively unfavorable and unmarked adjectives evaluatively favorable.

But, what happens when communicators describe themselves to others? While making self-descriptions communicators may wish to appear modest, and this desire may vary as a function of the culture and the person's societal role (e.g., Cialdini et al, 1998). Modesty can be achieved by using negations of unfavorable attributes (e.g., "my performance was not bad") instead of the corresponding favorable attributes ("my performance was good").

In order to ferret out the communicator's "true" sentiment, listeners must remove the "contamination" due to the normative constraints on communications – in our examples, the constraints imposed by the norms of politeness or modesty. When they suspect politeness, listeners should adjust the interpretation of a negation of a favorable [unmarked] attribute toward the antonym ("she is not pretty" means "she is ugly") to a greater extent than they should adjust the interpretation of a negated unfavorable [marked] adjective. However, the opposite effect should occur if listeners suspect modesty. The likelihood that a negation of an unfavorable attribute (e.g.,

“not bad”) would be understood as the opposite attribute (“good”) should be greater than the likelihood that negation of a favorable attribute (“not good”) would be understood as the opposite (“bad”). The context cues listeners to interpret negations differently in situations involving politeness and modesty.

It is easy to envision other situations in which one uses a negation rather than the alternative opposite term, even when such a term exists. Accordingly, we should distinguish between two kinds of unipolar concepts: *intrinsic* unipolar terms, for which there is no clear-cut opposite term, and *extrinsic* unipolar terms, which have opposite terms whose use is suppressed due to politeness or political correctness. I propose, speculatively, that the meaning of negations involving intrinsic unipolar concepts gravitate toward the unipolar concepts, while the meanings of negations of extrinsic unipolar concepts gravitate toward their opposites.

Final comments

Let me end by discussing several directions in which this research might be extended. To begin with, negation terms may differ in the amount of shift toward the opposite which they induce. As a reviewer of this chapter noted, negating an attribute by “un” is typically stronger than negating it by “not”. So if Tom is described as “unhappy” listeners might feel comfortable saying that he is “not happy”, but if Gary is described as “not happy” listeners might be reluctant to call him “unhappy”. This raises the possibility that “un” is more likely to be used with unmarked adjectives than with marked adjectives. I do not know of empirical evidence for this hypothesis.

A second, more remote extension of our research involves investigation of other qualifiers and quantifiers. How does “almost all” vs “almost none” interpreted when applied to marked and unmarked options? Moxey (2006; Sanford & Moxey, 2003) noted that the terms used as quantifiers might direct attention either to the set of exemplars mentioned in the sentence (which they termed positive quantifiers) or to the complementary set of exemplars (negative quantifiers). It was proposed that the latter type of quantifiers operate mentally by creating an expectation and denying

it. It might be interesting to explore whether the difference between positive and negative quantifiers interacts with the markedness of description, being weaker when the less expected (marked) term is used (cf., Sanford, Fay, Stewart, & Moxey, 2002).

Finally, it is important to emphasize that the discussion so far has involved explicit negations, in which an attribute was denied by linguistic terms such as “not”, “never”, “non”, etc. There is, however, a different type of acts of negation which is triggered spontaneously by decision-makers even when none of the explicit terms signaling negation is communicated. Assume that you hear a statement from a source you distrust, whether your political opponent, your next-door neighbor with whom you have a conflict of interest, or your disgruntled lover whom you abandoned two months ago. What might be common to these sources is that you trust that they are not trustworthy – they seem (so you believe) to twist the truth so as to make you look bad or feel bad. Imagine hearing one of these sources telling you that a particular movie is worth seeing, or that a specific individual is a nice person. Because you distrust them, it is likely that in processing their statements you will negate them. Such contextual negation might act differently than the explicit negation we have discussed so far because it is self-generated rather than explicit in the communication. Accordingly, future studies should examine whether contextual negations are quicker, drive the listener more toward the opposite, and persist longer. Speculatively, because contextual negations are hypothesized to lead the listener more toward the opposite, they might produce weaker framing effects than the standard explicit negation.

References

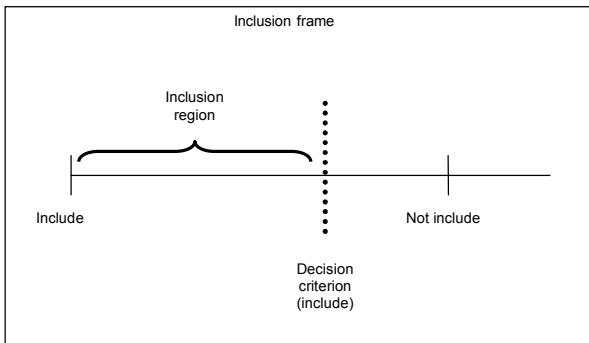
- Alia-Klein, N., Goldstein, R.Z., Tomasi, D., Zhang, L., Fagin-Jones, S. Telang, F. Wang, G., Fowler, J.S., & Volkow N.D. (2007) What is in a word? No versus Yes differentially engage the lateral orbitofrontal cortex. *Emotion*, 7, 649–659
- Battistella, E.L. (1996). *The logic of markedness*. NY, Oxford: Oxford University Press.
- Blackmore, S.J., and Decety, J. (2001) From the perception of action to the understanding of intention, *Nature Reviews Neuroscience*, 2, 561–567.
- Bonnefon, J., & Villejoubert G. (2006) Tactful or doubtful? Expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science*, 17, 747 - 751
- Boucher, T., & Osgood, C.E. (1969). The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior*, 8, 1-8.
- Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 82, 45–73.
- Cialdini, R.B., Wosinska, W., Dabul, A.J., Whetstone-Dion, R., & Heszen, I. (1998). When social role salience leads to social role rejection: Modest self-presentation among women and men in two cultures. *Personality and Social Psychology Bulletin*, 24, 473-481.
- Clark, H.H. & Clark, E.V. (1977). *Psychology and Language: An Introduction to Psycholinguistics*. Harcourt Brace Jovanovich, San Diego.
- Colston, H.L. (1999). “Not Good” is “Bad”, but “Not bad” is not “Good”: An analysis of three accounts of negation asymmetry. *Discourse processes*, 28, 237-256.
- Fiedler K. (2000). Illusory correlations: A simple associative algorithm provides a convergent account of seemingly divergent paradigms. *Review of General Psychology*, 4, 25-58
- Fischler, I., Bloom, P.A., Childers, D.G., Roucos, S.E., & Perry, N.W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, 20, 400-408.
- Fraenkel, T (2006) *The meaning of negated adjectives and the preservation of negation in communication*. Dissertation submitted to the Senate of the Hebrew University.
- Fraenkel, T., & Schul, Y. (2008). The meaning of negated adjectives. *Intercultural Pragmatics*, 5, 517-540
- Gannon, K.M., & Ostrom, T.M. (1996). How meaning is given to rating scales: the effects of response language on category activation. *Journal of experimental social psychology*, 32, 337-360.
- Giora, R. (2006) Anything negatives can do affirmatives can do just as well, except for some metaphors. *Journal of Pragmatics*, 38, 981-1014.
- Giora, R., Balban, N., Fein O., & Alkabetz, I. (2005). Negation as positivity in disguise. In Colston, H. L., & Katz, A. (Eds.), *Figurative Language Comprehension: Social and Cultural Influences* (pp. 233-258). NJ, Erlbaum, Hillsdale.
- Givon, T. (1978). Negation in language: Pragmatics, function, ontology. In Cole, P. (ed.), *Syntax and semantics 9: Pragmatics*. New York: Academic Press, pp. 69-112
- Grice, H.P. (1975). Logic and conversation. In P. Cole & J. Morgan (eds.), *Syntax and Semantics: Vol 3, Speech Acts* (pp. 43-58). New York: Academic Press.

- Hage, P. (2001) Marking theory and kinship analysis: Cross-cultural and historical applications. *Anthropological Theory*, 1, 197-211.
- Hartmann, R.R.K., & Stork, F.C. (1972). *Dictionary of language and linguistics*. London: Applied Science Publishers.
- Hasson, U., & Gluckberg, S. (2006) Does understanding negation entail affirmation? An examination of negated metaphors. *Journal of Pragmatics*, 38, 1015–1032
- Heider, F., & Simmel, M. (1944) An Experimental Study of Apparent Behavior. *The American Journal of Psychology*, 57, 243-259
- Holleman, B.C., and Pander Maat, H.L.W (2009) The pragmatics of profiling: Framing effects in text interpretation and text production. *Journal of Pragmatics*, 41, 2204-2221
- Horn, L.R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicatures. In D. Schiffrin (ed.), *Meaning, Form, and Use in Context* (pp. 11-42). Washington: Georgetown University Press.
- Horn, L.R. (1989). *A natural history of negation*. Chicago: Chicago university press.
- Horn, L. (2004) Implicature. In G. Ward and L. Horn (Eds.) *The handbook of pragmatics*. Blackwell Publishing (2004). (pp 3-28)
- Jordan, M.P. (1998). The power of negation in English: Text, context and relevance. *Journal of Pragmatics*, 29: 705-752.
- Just, M.A., & Clark, H.H. (1973). Drawing inferences from the presuppositions and implications of affirmative and negative sentences. *Journal of verbal learning and verbal behavior*, 12, 21- 31.
- Kahneman D., & Miller, D.T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93, 136-153.
- Kaup, B., Ludtke, J., & Zwaan, R.A. (2006) Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, 38, 1033-1050
- Ladusaw, W.A. (1996). Negation and polarity items. In S. Lappin (Ed.), *The handbook of contemporary semantic theory* (pp. 321-341). Oxford: Blackwell Publishers.
- Lea, R.B., & Mulligan, E.J. (2002). The effect of negation on deductive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 303–17.
- Lehrer, A. (1985). Markedness and antonymy. *Journal of Linguistics*, 21, 397-429.
- Levin, I. P., Jasper, J. D. & Forbes, W. S. (1998). Choosing versus rejecting options at different stages of decision making. *Journal of Behavioral Decision Making*, 11, 193-210.
- Levinson, S.C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- MacDonald, M.C., & Just, M.A. (1989). Changes in activation levels with negation. *Journal of Experimental Psychology*, 15, 633-642.
- Matlin, M.W. & Stang, D.J. (1978). *The Pollyanna principle: Selectivity in language, memory, and thought*. Cambridge, Mass.: Schenkman
- Mayo, R., Schul, Y., & Burnstein E. (2004). “I am guilty” vs. “I am innocent”: Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology*, 40, 443-449.

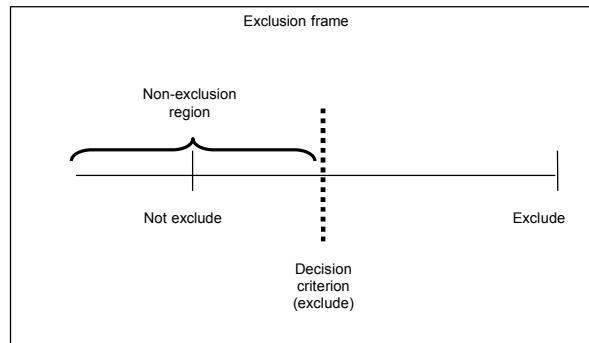
- Moxey, L.M. (2006) Effects of what is expected on the focusing properties of quantifiers: A test of the presupposition-denial account. *Journal of Memory and Language* 55, 422-439.
- Rozin, P., Berman, L., & Royzman, E. (in press) Biases in use of positive and negative words across twenty natural languages. *Cognition and Emotion*
- Sanford, A. J. & Moxey, L. M. (2003) New Perspectives on the Expression of Quantity *Current Directions in Psychological Science* 12, 240-243.
- Sanford, A. J., Fay, N., Stewart, A. J., & Moxey, L.M. (2002) Perspectives in statements of quantity, with implications for consumer psychology. *Psychological Science* 13, 130-134.
- Santos, T. (1987) Markedness theory and error evaluation: An experimental study. *Applied Linguistics*, 8, 207-218
- Schul, Y., & Burnstein, E. (1985). When discounting fails: Conditions under which individuals use discredited information in making a judgment. *Journal of Personality and Social Psychology*, 49, 894-903
- Schul, Y., & Burnstein, E. (1998). Suspicion and discounting: Ignoring invalid information in an uncertain environment. In J.M. Golding & C. MacLeod (Eds.) *Intentional forgetting: Interdisciplinary approaches*. Erlbaum. (pp. 321-348).
- Schul, Y., & Mayo, R. (1999) Two sources are better than one: The effects of ignoring one message on using a different message from the same source. *Journal of Experimental Social Psychology*, 35, 327-345.
- Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition*, 101, 467-494.
- Tversky A., & Kahneman, D. (1981) The framing of decisions and the psychology of choice. *Science*, 211, 453-458.
- Wason, P.C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, 4, 7-11.
- Whitmeyer, J.M. (1997). Contrasting trait adjectives as distinct psychological entities. *Psychological reports*, 80, 739-743.
- Yaniv, I., & Schul, Y. (1997). Elimination and inclusion procedures in judgment under uncertainty. *Journal of Behavioral Decision Making*, 10, 211-220
- Yaniv, I., & Schul, Y. (2000) Acceptance and elimination procedures in choice: Non-complementarity and the role of implied status quo. *Organizational Behavior and Human Decision Processes*, 82, 293-313.
- Yorke, M. (2001). Bipolarity... or not? Some Conceptual Problems Relating to Bipolar Rating Scales. *British Educational Research Journal*, 27, 171-186.

Figure Caption

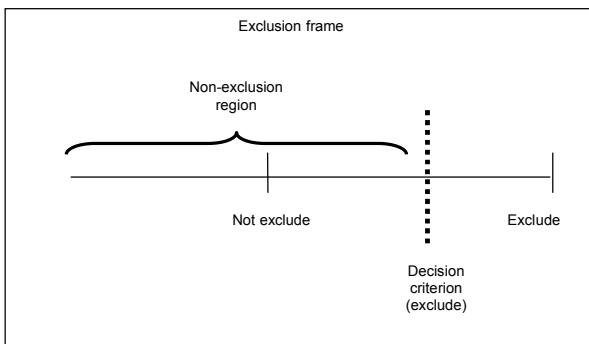
Figure 1: Schematic depiction of the alternative decision problems under inclusion and exclusion frames of mind.



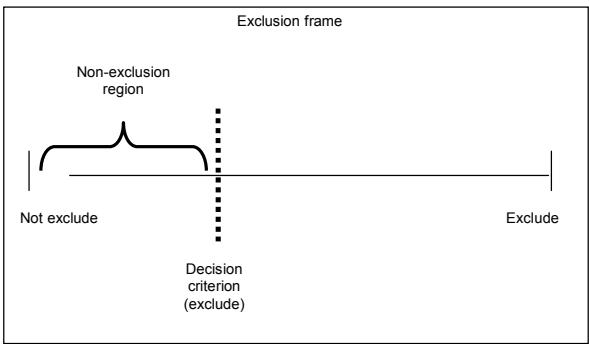
(a)



(b)



(c)



(d)