Chapter 8
Framing

Till Grüne-Yanoff

Abstract The concept of framing, experimental evidence supporting framing effects, and models and theories of decision-making sensitive to framing play important roles in policy analysis. First, they are used to caution about various elements of uncertainty that are introduced through framing into policy interventions. Second, framing is often referred to in order to justify certain policy interventions, as framing effects are often seen as sources of irrationality in need of correction. Third, framing effects are often used as instruments for policy-making, as they are seen as effective ways to influence behaviour. This review discusses the different concepts of framing, surveys some of the experimental evidence, described the dominant descriptive theories and the main attempts to assess the rationality or irrationality of behaviour sensitive to framing in order to clarify how exactly framing is relevant for policy making.

Keywords Framing • Preferences • Lotteries • Uncertainty • Behavioural economics • Mechanisms • Descriptive decision theory • Normative decision theory • Bounded rationality • Behavioural policy • Nudge • Boost

1 Introduction

There are usually many different ways in which we can frame a decision. This chapter clarifies what is meant by framing, why it is important for decision-making and how we can argue rationally about the choice of frames. Specifically, I briefly survey the history of the technical term in psychology (Sect. 2) and then illustrate the use of the term at the hand of various experimental studies in psychology and economics (Sect. 3). Sections 4 and 5 survey attempts to produce descriptively adequate accounts of the thus elicited phenomena, in terms of mechanistic models and more abstract theory, respectively. Section 6 focuses on the philosophical discussion to what extent framing phenomena are irrational, and why they should or should not be. Section 7 discusses some normative theories of framing, which

T. Grüne-Yanoff (✉)
Royal Institute of Technology (KTH), Stockholm, Sweden
e-mail: gryne@kth.se

S.O. Hansson, G. Hirsch Hadorn (eds.), *The Argumentative Turn in Policy Analysis*, Logic, Argumentation & Reasoning 10, DOI 10.1007/978-3-319-30549-3_8

seek to provide some room for rational choice being influenced by frames, and at the same time impose constraints on what "rationally framed" decisions could be. Section 8, finally, addresses how the scientific discussion of framing has led to different policy proposals how to mitigate framing effects, and how framing effects should be used to influence people's decision.

Framing relates to uncertainty in multiple ways. First, the effect of framing on decisions is often observed in contexts involving uncertainty. For example, it matters sometimes whether an uncertain outcome is differentiated into some very unlikely events and some more likely outcomes, or whether this outcome is described as one bundle with a mean probability of all its events. Second, frames also create uncertainty, for example with respect to an individual's preferences. If an agent changes preferences over options under seemly irrelevant changes of the framing, the uncertainty about that individual's preferences (their authenticity, or their relevance for welfare properties) increases. Furthermore, the fact that frames affect decisions also creates uncertainty about the rationality of these decisions: they might be unduly influenced by these frames, and alternative ways how to arrive at these decisions might be required instead. Overall, these considerations provide arguments against an algorithmic perspective on decision-making (see Hansson and Hirsch Hadorn 2016). Such an algorithmic perspective claims that with sufficient information, decision-making consists in the application of a fully specified procedure (an algorithm), which yields an unambiguous outcome. Contrary to that, framing yields uncertainties that limit the straightforward application of algorithms. Furthermore, deliberation requires reconstruction and analysis of different framings of a decision problem, and this is a task of argumentative methods, not mere application of algorithms (see Brun and Betz 2016). Hence, considerations of framing support the argumentative turn of policymaking.

2 History and Taxonomy of the Term "Framing"

In the context of decision theory, Tversky and Kahneman (1981) were the first to propose the term "framing". They define a "decision frame" as:

> the decision maker's conception of the acts, outcomes, and contingencies associated with a particular choice. . .controlled partly by the formulation of the problem, and partly by the norms, habits, and personal characteristics of the decision maker. (Tversky and Kahneman 1981:453)

Crucial for the understanding of decision framing is the claim that one and the same element of a decision problem, when considered from different frames, might appear in different ways, and these appearances might be decision-relevant. For example, a glass can be described either as half-full or as half-empty, and people might consider these two descriptions of the same outcome as the descriptions of two different outcomes. Similarly, a body movement like forming a fist can be described a single act, or as the sequence of movements that constitute that act.

Finally, the relevant future states of the world can be described in more or less detail. When describing tomorrow's possible states of the weather, for example, I might distinguish (i) only "sunshine" or "no sunshine" or I might distinguish (ii) "sunshine", "clouds", "rain", or "snow". Framing in the wide sense refers to the fact that in order to analyse a decision, one always needs to delineate a decision problem or embed it in a particular context (see Doorn 2016; Elliott 2016; Grunwald 2016). This is of course related to a more general attitude towards or thinking about the world (e.g. Goffman 1974), as for example expressed in various forms of discourse analysis. Framing in the narrower sense only concerns how the conception (description and structuring) of the specific decision problem has an effect on decision-making. Of course, because this effect is often not known in advance, the wide and the narrow notion of framing are sometimes not clearly separated.

To distinguish framing with respect to what is framed, Tversky and Kahneman (1981) characterize three kinds of framing:

(A) framing of outcomes,
(B) framing of acts, and
(C) framing of contingencies.

Of these three types, framing of outcomes has received most attention in the literature and is the form most closely associated with the term "framing." As in the glass half-full/half-empty example, outcome framing is typically taken to affect the decision maker's evaluation of the outcome. Therefore, this type is also known as "valence framing" (Levin et al. 1998), which often is differentiated into three sub-types:

(A1) risky choice framing
(A2) attribute framing
(A3) goal framing

Risky choice framing is performed by re-describing the consequences of risky prospects, for example by re-describing a 70 % post-surgery survival chance as a 30 % chance of dying from this surgery. Tversky and Kahneman seem to be the first to describe this type. *Attribute framing* is achieved by re-describing one attribute of the objects to be evaluated, for example by re-describing a glass that is half-full as a glass that is half-empty. This type of framing has been investigated before Tversky and Kahneman, for example by Thaler (1980). *Goal framing*, finally, consists not in a re-description of the outcome directly, but rather in a re-description of the goal by which outcomes are evaluated. For example, one can evaluate monetary outcomes of one's acts either with the goal of "maximizing wealth" or with the goal of "avoiding any unnecessary losses". A re-description is different from a revision of the goal (see Edvardsson Björnberg 2016).

The types of framing discussed so far all concern the conception of a decision problem "controlled ... by the formulation of the problem", as Tversky and Kahneman put it in the above quotation. Here framing is constituted by the

description or re-description of elements of a decision problem. Partly because this description-factor can be experimentally manipulated with relative ease, most of the literature has focused on these types (as will become clear in the description of the different experimental designs used). However, framing is not restricted to this, as Tversky and Kahneman themselves acknowledge: framing is affected "partly by the norms, habits, and personal characteristics of the decision maker" (ibid.). Kühberger (1998) stresses this aspect of framing when he distinguishes between a "strict" and in a "loose" sense of the framing concept. The strict sense corresponds to those types of framing that are affected by redescription. The loose definition, however,

> refers to framing as an internal event that can be induced not only by semantic manipulations but may result also from other contextual features of a situation and from individual factors, provided that problems are equivalent from the perspective of economic theory. Describing equivalent dilemmata as a give-some vs. as a take-some dilemma is an example of this type of framing. (Kühberger 1998:24)

This introduces elements of the wide sense of framing back into the picture: any delineation and structuring of the decision problem might have an effect on decision-making, even if these are hard to categorise with the tools of decision theory. Unsurprisingly, such cases have been far less discussed in the literature. The following taxonomy therefore cannot be considered comprehensive. Nevertheless, the following distinctions might be useful:

(D) Procedural framing
(E) Ethically loaded frames
(F) Temporal frames

Gold and List (2004) argue that the ways how mental attitudes are elicited or measured constitutes procedural framing. For example, Lichtenstein and Slovic (1971) devised different ways how to elicit people's preferences over the same prospects. They found that the elicited preferences strongly depended on the elicitation procedure, up to the point where the differently elicited preferences over the same prospects became inconsistent. Gold and List therefore argue that such elicitation procedures constitute a kind of framing.

In social dilemma and coordination games, Bacharach et al. (2006) identify different ethically loaded frames that a player may adopt, namely the *I-frame* and the *we-frame*. Standard game theory implicitly assumes that a player in cases like the Prisoners' Dilemma always adopts an I-frame (asking "What should I do?"), leading to the dominant reasoning ("whatever others do, I will be better off defecting"). But she could be adopting, argue Bacharach et al. (2006), a we-frame (asking "What should we do?"). Players who adopted a we-frame will choose to cooperate in social dilemmas, as this contributes to the strategy profile that maximizes the group's payoff. Bacharach explicitly calls such cases "framing"; research on these phenomena, however predates the framing terminology (e.g. Evans and Crumbaugh 1966). Some authors seek to subsume ethically loaded frames under goal framing (Levin et al. 1998:168).

Tversky and Kahneman (1981) briefly mention another kind of framing, namely the changing of temporal perspectives.

> The metaphor of changing perspective can be applied to other phenomena of choice, in addition to the framing effects with which we have been concerned here. The problem of self-control is naturally construed in these terms. . ..an action taken in the present renders inoperative an anticipated future preference. An unusual feature of the problem of intertemporal conflict is that the agent who views a problem from a particular temporal perspective is also aware of the conflicting views that future perspectives will offer. (Tversky and Kahneman 1981:457)

In cases of intertemporal conflict – for example doing things now or later – a decision maker can assume the perspectives of her different temporal selves. Assuming today's perspective will let the decision maker decide according to her current preferences, while assuming her future self's perspective will give her future preferences an influence (see Hirsch Hadorn 2016; Möller 2016). Tversky and Kahneman seem to suggest that these perspectives correspond to different temporal frames, although this language has not been widely adopted in the literature.

Clearly, other applications of framing in this loose sense are possible, but because they are not widespread in the literature, I will not discuss them here. Instead, I will briefly sketch three motivations that led Tversky and Kahneman to introduce the concept, and that contributed to its pervasive adoption in the literature.

First, before the presentation of the framing concept in 1981, Tversky and Kahneman had developed a new research paradigm in psychology, that sought to document systematic deviations of experimental subjects from the prediction of the standard rational choice model (Heukelom 2014). The experimental elicitation of framing phenomena stands in this tradition, as standard rational choice models descriptively and normatively assume that people's decisions are invariant under alternative descriptions of the same decision elements (I will discuss the normative assumption of these standard models for randomized controlled trials (RCTs) more in Sect. 6). As part of this broader research effort, other researchers experimentally investigated behaviour that conceptually is very close to framing, although they did not use this terminology (e.g.; Thaler 1980; Lichtenstein and Slovic 1971).

Second, Kahneman and Tversky (1979) famously proposed "prospect theory" in order to model the systematic deviations that they and other researchers had elicited. Although there is no terminological reference to framing in prospect theory, the theory relies on evidence that conceptually is very close to cases of valence framing. Unsurprisingly, Tversky and Kahneman (1981) then propose prospect theory as an explanation of the framing effects they describe.

Third, many researchers who seized on the framing concept, including Tversky and Kahneman, claim it as a model for understanding anomalous economic phenomena in the real world that cannot be explained with standard economic models. Kahneman and Tversky (1984:347), for example, claim that framing is the factor underlying the observation "that the standard deviation of the prices that different stores in a city quote for the same product is roughly proportional to the average

price of that product (Pratt et al. 1979).” Bacharach (2001:4) argues that framing lies at the bottom of the “Money illusion”, and Kahneman and Tversky (1984:349) argue that observations of inconsistent choices of gambles and insurance policies (as described e.g. by Hershey and Schoemaker 1980) are driven by framing.

To conclude this section, I would like to point out a certain tension in the research on framing. On the one hand, sustained research activity has produced a manifold of experimental designs (surveyed in Sect. 3) and mechanistic models (Sect. 4). These findings correspond well with the multitude of framing concepts that I discussed in this section, and which seem to suggest that framing should not be treated as a very unified concept. On the other hand, however, the continued use of the term ‘framing’ for all these seemingly diverse concepts suggests that its users see a deeper unity in the concept of framing. On an abstract level, all these concepts are seen as closely interlinked. As Bacharach put it: “A frame is the set of concepts or predicates an agent uses in thinking about the world. . . One does not just see, but one sees as” (Bacharach 2001:1). This has given rise to a tendency to seek unified theories of framing (as discussed in Sects. 5 and 7) and derive general claims about when framing effects justify policy interventions or which framing effects can be exploited for policy purposes. One of the purposes of this review is to represent this tension and its determinants appropriately, which hopefully might contribute to its solution.

3 Experimental Elicitation of Framing Phenomena

Framing is fundamentally an experimentally identified phenomenon. Only the presentation of re-described acts, states or outcomes under highly controlled conditions have yielded behavioural evidence for the systematic deviation from standard RCT models. Because of this strong dependence on experiments, understanding the concept (or the concepts) of framing requires looking into the details of the experiments that elicited this behavioural evidence.

Many hundreds of experimental studies on framing have been published since 1981. It is not the purpose of this section to provide a systematic review of these. The interested reader might instead consult extant reviews (Levin et al. 1998) and meta-analyses (Gallagher and Updegraff 2012; Gambara and Pinon 2005; Kühberger 1998). The overall tenor of these is that the framing effect is a robust phenomenon:

> A meta-analysis of 136 research reports yielded 230 single effect sizes, which, overall, corroborated the framing effect. (Kühberger 1998:47)

Yet this conclusion disguises an important heterogeneity. Not only do such meta-analyses draw on substantially different experimental designs, they also disclose a heterogeneity of effect sizes, depending on the respective experimental designs. I will come back to this at the end of this section. First, I will describe some experiment types, in order to make obvious the heterogeneity in design.

Tversky and Kahneman's (1981) "Asian disease problem" is clearly the prototypical and most-cited example of a framing experiment. They presented two separate groups of experimental subjects with one of the following decision problems. Number of participants and response frequencies are described in rectangular brackets (Tversky and Kahneman 1981:453):

> *Problem 1* [N = 152]: Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:
>
> - If Program A is adopted, 200 people will be saved [72 percent]
> - 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. [28 percent]
>
> Which of the two programs would you favor?
> *Problem 2* [N = 155]:
>
> - If Program C is adopted 400 people will die. [22 percent]
> - If Program D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die. [78 percent]
>
> Which of the two programs would you favor?

The experiment poses two discrete choices between a risky and a riskless option of equal expected value. In one problem, the options are described in positive terms (i.e., lives saved); in the other in negative terms (i.e., lives lost). Because the experimental manipulation consists in a re-description of a consequence of a risky choice, this is a framing of type (A1), as described in the previous section.

Tversky and Kahneman observed a "choice reversal," where the majority of subjects who were given the positively framed problem chose the option with the certain outcome, whereas the majority of subjects who were given the negatively framed problem chose the risky option.

Despite its prototypical status, following framing experiments have often deviated substantially from the Asian disease design. This has led some authors to question whether these experiments provide evidence for the same phenomenon:

> many recent studies of valence framing effects have deviated greatly from the operational definitions and theoretical concepts used in the original studies, thus stretching the limits of Kahneman and Tversky's initial theoretical accounts. (Levin et al. 1998:151)
>
> Diverse operational, methodical and task-specific features make the body of data heterogeneous to a degree that makes it impossible to speak of 'the framing effect.' (Kühberger 1998:43)

To make these worries more salient, let me summarize some of the main differences in experimental designs (in this I largely follow Kühberger 1998:32–33). The first difference concerns the nature of the options. In some experimental designs, one option is riskless and the other is risky – for example in the Asian disease design described above. In others, both options are risky, as for examples when subjects are asked to choose between therapies that are risky to different degrees. The second difference concerns the degree of partitioning of risky option. In many designs, each risky option only consists of a dual partition,

with an event either occurring or not occurring. In other designs, for example bargaining tasks, options might be partitioned more finely. A third difference concerns the nature of the framing manipulation. Framing can be manipulated either by explicit labelling (e.g. "win" vs. "loose"; "gain" v. "pay") or by implicitly describing the task in value-relevant ways (e.g. by describing a situation either as a commons-dilemma or a public goods problem). A fourth difference concerns the subjects' responses: they might be asked to choose between options, as in the Asian disease design, or only to rank the different options. A fifth difference between designs concerns the comparison of choices: are choices of the same person in the two different situations compared, or are the compared choices those of different people (as in the Asian disease problem)? Finally, designs vary in the domain of their choices, involving either economic, social, medical or gambling decisions. Thus, the design of experiments that all are supposed to provide evidence for or against framing effects substantially differs.

Furthermore, framing phenomena have also been elicited in inferential tasks, which do not involve the choice between acts, but rather the choice of theoretical conclusions. Many studies in this area have concluded that laypeople and professionals alike (see Koehler 1996; Berwick et al. 1981) make poor diagnostic inferences on the basis of statistical information. In particular, their statistical inferences do not follow Bayes' theorem—a finding that prompted Kahneman and Tversky (1972:450) to conclude: "In his evaluation of evidence, man is apparently not a conservative Bayesian: he is not Bayesian at all." The studies from which this and similar conclusions were drawn presented information in the form of probabilities and percentages. From a mathematical viewpoint, it is irrelevant whether statistical information is presented in probabilities, percentages, absolute frequencies, or some other form, because these different representations can be mapped onto one another in a one-to-one fashion. Seen from a psychological viewpoint, however, as the proponents of the boost approach have argued, representation does matter: Some representations make people more competent to reason in a Bayesian way in the absence of any explicit instruction (Hoffrage et al. 2000; Gigerenzer and Hoffrage 1995).

That the experimental designs for the elicitation of framing differ substantially perhaps would not be a problem if these designs all yielded comparable effects – indeed, such a result would even support the robustness of the framing effect. Unfortunately, this does not seem to be the case. Rather, effect sizes obtained from different experimental designs systematically differ:

> The more experiments differ from the original Asian disease problem, the lesser the reference point effect. . .. Overall, 4 of 10 procedural designs are ineffective: the Clinical reasoning design is ineffective, and, to make things worse, is used relatively frequently. Further ineffective designs are Escalation of commitment, Message compliance, and Evaluation of objects. (Kühberger 1998:45)
>
> the likelihood of obtaining choice reversals was directly related to the similarity between features of a given study and features of Tversky and Kahneman's (1981) original 'Asian disease problem.' (Levin et al. 1998:157)

This of course does not invalidate the framing concept altogether, but it should caution against its context-free use: the phenomenon of framing in some important way depends on the design of the manipulation and the environment in which it is elicited. Because the determining factors of this elicitation are not yet fully understood, it is difficult to extrapolate from the laboratory conditions to other contexts. To progress in this matter would require knowing more about the underlying mechanisms through which these environmental factors influence framing (Grüne-Yanoff 2015). I will discuss this topic in the next section.

4 Possible Mechanisms of Framing

Evidence for framing phenomena typically comes in the form of effect sizes – a qualitative measure of the correlation between framing manipulation and behavioural changes. These relations are captured by some of the theories discussed in Sect. 5. What remains often opaque is the process through which the framing produces the change.

> Cognitive processes are another stepchild of framing research. Taken the effect for granted (what can safely be assumed), we would be well advised to probe for the cognitive processes and structures that are responsible for it. (Kühberger 1998:47)

This is of particular relevance given the heterogeneity of effect sizes and their seeming dependence on experimental design. One possible explanation for this dependence is that different framing manipulations in different circumstances trigger different cognitive mechanisms, which then consequently produce different effects and different effect sizes.

There is very little research on the cognitive mechanisms underlying framing. Mechanisms typically only appear as mere speculations and ad-hoc how-possibly explanations of observed phenomena. Nevertheless, it is informative to discuss some of these speculations in order to gain an understanding of their diversity.

For the framing of outcomes, for example, Tversky and Kahneman propose *contextual referencing* as a cognitive mechanism:

> There are situations, however, in which the outcomes of an act affect the balance in an account that was previously set up by a related act. In these cases, the decision at hand may be evaluated in terms of a more inclusive account, as in the case of the bettor who views the last race in the context of earlier losses. (Tversky and Kahneman 1981:457)

For the framing of contingencies, multiple cognitive processes have been proposed. For example, Tversky and Kahneman (1981) propose a *pseudocertainty effect*, which consist of an illusionary of certainty. Options that are certain, they suggest, are preferred to options that are uncertain. If now an uncertain option is divided into two sequential steps, one of which incorporates all uncertainty, then the decision maker might take the appearance of certainty from the second step as relevant for the whole option, and prefer it as if it were certain.

Another possible cognitive mechanism behind the framing of contingencies might be *limited imagination*:

> the fundamental problem of probability assessment [is perhaps] the need to consider unavailable possibilities. . .People. . .cannot be expected. . .to generate all relevant future scenarios. Tversky and Koehler (1994:565)

That is, because people are unable to imagine relevant possible scenarios, they do not partition contingencies finely enough. But when they are given such scenarios from external sources, they incorporate them into the decision problem and decide accordingly, thus leading to framing effects.

A further possible cognitive mechanism behind the framing of contingencies might be *limited memory*. Even if they have already heard about possible contingencies, they might have forgotten about them again. Provision of more detailed descriptions then might help in remembering such contingencies (and their relevance), leading to framing effects.

Yet another possible mechanism of framing effects is that different descriptions alter the *salience* of events. For example, by re-describing a week either as a single event or as a sequence of 7 days, Fox and Rottenstreich (2003) elicited substantially different answers from subjects asked to report the probability that Sunday would be the hottest day of the coming week. In such cases, descriptions produce framing effects without fostering imagination or recall.

5 Descriptive Theories of Framing

Despite the diversity in concepts, elicitations and mechanisms of framing, various general theories of behaviour have been proposed that claim to adequately describe the framing phenomenon. None of these theories have mechanistic or procedural content; rather, they aim to capture the systematic relationship between framing manipulation and behavioural changes only. This section briefly reviews four such attempts, namely Prospect Theory, Cumulative Prospect Theory, Support Theory and Partition-dependent Expected Utility Theory. Notably, these theories seek to *describe* actual behaviour, influence, amongst other factors, by framing, while refraining to judge whether this behaviour is rational or not.

Prospect theory (Kahneman and Tversky 1979) describes behaviour as influenced by the decision maker's evaluation that is generated relative to a certain reference point. The theory proposes a two-step decision process: in the editing phase, a reference point is set. In the evaluation phase, outcomes are evaluated either as gains or losses, relative to the set reference point. Specifically, people evaluate gains (i.e. outcomes above the reference point) differently than losses (i.e. outcomes below the reference point) and care generally more about potential losses than potential gains. Prospect theory predates the explicit conceptualization of framing, but it clearly captures its main idea: namely, that the

presentation of the outcomes of a decision problem systematically influences the decision maker's choice. That the glass is half-full rather than half-empty makes sense only against changing reference points – people consider it half-empty if their reference point was (the expectation of) a full glass, while they consider it half-full if their reference point was an empty glass. Similar with outcomes of medical interventions that are described either as a chance of death or of survival – people will focus more on the chance of death caused by a medical intervention if their reference point is the certain expectation of surviving, while they focus more on the chance of survival if their reference point is the certain expectation of dying.

In 1992, Tversky and Kahneman proposed a new theory, *cumulative prospect theory*, replacing the 1979 approach. In the new model, the editing phase of prospect theory was renamed "framing phase" (Tversky and Kahneman 1992). Furthermore, people tend to overweight extreme yet unlikely events, but underweight "average" events. The main difference to Prospect Theory is that cumulative probabilities are transformed, rather than the probabilities themselves. Cumulative prospect theory, with the new framing phase and the focus on cumulative probabilities, no longer implies violation of stochastic dominance and makes the generalization to arbitrary outcome distributions easier. It is therefore on theoretical grounds an improvement over Prospect Theory.

While the above versions of prospect theory describe evaluations of outcomes as dependent on reference points, hence focusing on framing of outcomes, the following theories focus on the framing of contingencies and acts. Tversky and Koehler's (1994) *support theory* describes how probability judgments are affected by whether propositions are presented as explicit or implicit disjunctions. For example, subjects are asked to judge how probable it is that a randomly selected person "will die from an accident". Subjects tend to give a lower probability to this implicit conjunction, than they give to an explicit conjunction consisting of "a randomly selected person will die from a car crash", "... a plane crash", "...a fire", "... drowning", etc. Support theory accounts for this phenomenon by describing agents as assigning subjective probability to hypotheses. Subjective probability increases as hypotheses are "unpacked" into more explicit disjunctions. Specifically, while probabilities are complementary in the binary case, they are subadditive in the general case.

Ahn and Ergin's (2010) *partition-dependent expected utility theory* allows discriminating between different presentations of the same act. Starting from the standard subjective model of decision-making under uncertainty, they distinguish different expressions for an act as distinct choice objects. Specifically, lists of contingencies with associated outcomes are taken as the primitive objects of choice. Choices over lists are represented by a family of preferences, where each preference is indexed by a partition of the state space. The respective partitions are interpreted as descriptions of the different events.

6 Normative Assessment of Framing

The concept of framing is inextricably linked to normative judgment. Tversky and Kahneman argued that framing leads to preference reversals, violating consistency requirements of standard decision theory:

> we describe decision problems in which people systematically violate the requirements of consistency and coherence (Tversky and Kahneman 1981:453)

Upon closer inspection, however, it isn't entirely obvious which consistency requirements of standard decision theory framing supposedly violates. None of the axiomatisations of von Neumann and Morgenstern (1944), Savage (1954), Anscombe and Aumann (1963) or Jeffrey (1963) contain any explicitly formulated axiom that the standard framing cases would violate.[1]

Instead, the formulation of the framing effect led to the explicit formulation of a rationality axiom that previously had been implicitly assumed. This requirement has been variably called the *principle of invariance* or the *principle of extensionality*. Kahneman and Tversky formulate it thus:

> Invariance requires that the preference order between prospects should not depend on the manner in which they are described. In particular, two versions of a choice problem that are recognized to be equivalent when shown together should elicit the same preference even when shown separately. (Kahneman and Tversky 1984:343)

Arrow formulated the principle of extensionality thus:

> A fundamental element of rationality, so elementary that we hardly notice it, is, in logicians' language, its extensionality. The chosen element depends on the opportunity set from which the choice is to be made, independently of how that set is described (Arrow 1982:6)

Arrow makes explicit reference to extensionality as a principle of logic. In logic, the principle of extensionality requires of two formulas that have the same truth-value under any truth assignment to be mutually substitutable *salva veritate* in a sentence that contains one of these formulas. Thus, "the glass is half-full" and "the glass is half-empty" have the same truth-value in all possible worlds, because they refer to the same fact of the matter. An agent whose choice is affected by how this same fact is described violates extensionality. In the following discussion, I will reserve *extensionality* as the principle based on logical equivalence in this sense; it is determined by the semantic characteristics of the explicit formulations only. In contrast, I will be using *invariance* for the principle based on non-logical versions

[1] A qualification is necessary here. Kahneman and Tversky for example argue that specific kinds of act-framing violate the principle of dominance: "the susceptibility to framing and the S-shaped value function produce a violation of dominance in a set of concurrent decisions" (Kahneman and Tversky 1984:344). Clearly, dominance is an explicitly formulated requirement in these standard axiomatisations. However, because only special cases of framing violate dominance, and because the normative judgment apparently goes beyond these cases, it cannot be dominance violation that lies at the basis of judging framing to be irrational.

of equivalence; it is determined by implicit suggestions, that trigger pragmatic inferences, e.g. on expectations. So, two different formulations are invariant, if they implicitly suggest the same pragmatic inferences.

Thus defined, the two principles differ substantially: two descriptions might be semantically identical and yet differ pragmatically – I will discuss an example later in this section. However, two descriptions might be pragmatically identical and yet differ semantically – for example when the semantic differences are pragmatically irrelevant. That this distinction is relevant will (hopefully) become clear in this section. Unfortunately, the distinction isn't always so clear in the literature. Because the extensionality principle is the much clearer concept, I will discuss its relation to rationality first, and then focus on the invariance principle later.

Tversky and Kahneman (1986) considered invariance (here understood as extensionality) as a tacit axiom of rationality:

> This principle of invariance is so basic that it is tacitly assumed in the characterization of options rather than explicitly stated as a testable axiom. (Tversky and Kahneman 1986: S253)

Indeed, it has been formally shown recently that Jeffrey-Bolker decision theory (Jeffrey 1963) contains extensionality as an implicit axiom (Bourgeois-Gironde and Giraud 2009:391). For explicit formulations of this axiom, see e.g. Rubinstein (2000), and Le Menestreland and Van Wassenhove (2001).

Given the either implicit or explicit assumption of extensionality in most accepted normative decision theories, framing phenomena seem to be clear violations of rationality:

> The failure of invariance is both pervasive and robust. It is as common among sophisticated respondents as among naive ones, and it is not eliminated even when the same respondents answer both questions within a few minutes. . . .In their stubborn appeal, framing effects resemble perceptual illusions more than computational errors. . .. The moral of these results is disturbing: Invariance is normatively essential, intuitively compelling, and psychologically unfeasible. (Kahneman and Tversky 1984:343–4)

Those, like Tversky and Kahneman, who consider the extensionality normatively necessary, but who see its violation as pervasive, distinguish between normatively valid theories of decision making – which adhere to the invariance principle – and descriptively adequate theories of decision making – which describe the ways how people systematically violate extensionality. Theories of the first kind include von Neumann and Morgenstern (1944), Savage (1954), Anscombe and Aumann (1963) or Jeffrey (1963), while theories of the second kind were described in Sect. 5.

However, is the principle of extensionality really a defensible rationality requirement? This question really has two parts. The first concerns extensionality as a requirement for *full rationality*. The second concerns whether some violations are *compatible* with a normatively valid model of *bounded rationality*. In the remainder of this section, I will discuss some criticisms of the validity of extensionality as a requirement of full rationality. In the next section, I will review some normative theories of bounded rationality that allow limited violations of invariance.

Tversky and Kahneman early on acknowledged that cognitive effort considerations might mitigate the irrationality of framing effects:

> These observations do not imply that preference reversals [arising from framing] are necessarily irrational. Like other intellectual limitations, discussed by Simon under the heading of 'bounded rationality,' the practice of acting on the most readily available frame can sometimes be justified by reference to the mental effort required to explore alternative frames and avoid potential inconsistencies. (Tversky and Kahneman 1981:458)

However, this argument relies on a contested narrow interpretation of Simon's concept of bounded rationality (Gigerenzer and Brighton 2009). Tversky and Kahneman in the above quote clearly consider the validity of bounded rationality models to depend on an accuracy-cost trade-off: not-too-catastrophic inconsistencies are justifiable if the costs of avoiding them would be unreasonably high. In contrast, Gigerenzer and Brighton argue that the validity of bounded rationality models depends on the reliability of the models in performing well for their designated tasks in the designated environments.

In the context of framing, we find such arguments at various places. For example, Sher and McKenzie (2006) argue that the framing of an outcome encodes relevant additional information, which most people intuitively understand. They show experimentally that subjects systematically distinguish between "half-full" and "half-empty" glasses. A full glass of water (A) and an empty one (B) are put on the table. The experimenter asks the participant to pour half of the water into the other glass, and then to place the "half-empty glass" at the edge of the table. Most people choose glass A, the previously full glass.

Such violations of extensionality are rational responses when the goal is e.g. to avoid regret, because the different descriptions of the same fact might convey different information about the expectations of the chooser. In the glass example, if the glass was originally full, the resultant regret from obtaining one-half the water is different from the case where the glass was originally empty. Note that distinguishing between "half-full" and "half-empty" glasses violates extensionality, because the semantic properties of any sentence remains unaffected when one replaces one formulation with the other. Instead, the relevant information is obtained through pragmatic inferences, not logical ones.

Such pragmatic inferences often depend on surprising detail. For example, it seems that incomplete specifications are often interpreted as implicit recommendations. In the Asian disease case, described in Sect. 3, the riskless options are not fully specified, stressing only the amount of survivors or fatalities, respectively. When researchers completely specified the riskless options, the framing effect in the Asian disease problem disappeared (Mandel 2001; Kühberger 1995). If subjects interpret incomplete specification as implicit recommendations, then again, it is perfectly rational for them to take this additional information into account.

Another argument against the necessity of extensionality as a rationality criterion comes from the observation of people's ability to solve coordination problems by exploiting 'focal points'. Bacharach (2001) provides a game-theoretic analysis of such coordination problems, in which players have to coordinate on one out of

many possible equilibria. This, Bacharach argues, depends on players being able to identify one strategy profile as 'focal'. In a problem where to meet in a big town, such a focal point might be the most notable monument of that town; in a problem when to simultaneously perform a certain action, such a focal point might be 12'o clock at noon; in a problem to independently choose the same number between 0 and 100, such a focal point might be 0, or 50, or 100. It is an empirical fact that people often are able to solve such coordination problems, without being able to communicate with each other. Instead, they exploit the fact that within a particular way of describing a town, the time or a numerical interval, certain elements "stick out": these elements appear more salient than others under that description, and consequently draw the players focus onto themselves. Of course such salience varies with the descriptive frame – it is for this reason that Bacharach identifies the violation of extensionality as a success condition for coordination on focal points:

> Human framing propensities stand behind the well-known ability of people to solve coordination problems by exploiting 'focal points'. Ironically, it is precisely their incompleteness that we can thank for this. . ..The partiality and instability of frames or 'conceptual boundedness' disables human agents in certain tasks — in particular, it makes them manipulable by framers. However, the sharedness of frames enables them to do well in other tasks, and in some cases it is important for this that the shared frame is partial. (Bacharach 2001:7–9)

The first lesson to learn from these arguments is that the rationality of framing effects cannot be decided on a logical principle of extensionality. In decision-theoretic contexts, it is not relevant whether alternative descriptions are semantically equivalent (i.e. whether they have the same truth-value in all possible worlds), but rather whether they are informationally equivalent. In the above two cases, different frames of decision problems, although semantically equivalent, carried different decision-relevant information with them, and therefore it was rational for the agents to choose differently under these different frames. Sher and McKenzie (2006), for example, separate the issue of informational relevance from that of extensionality:

> There is no normative problem with logically equivalent but information non-equivalent descriptions leading to different decisions. (Sher and McKenzie 2006:487)

To the contrary, rational agents should be indifferent between two co-reportive propositions if and only if the frames in which their common reference is expressed convey exactly the same information about choice-relevant pieces of information.

While this rejects the logical notion of extensionality as a rationality criterion for decision making, it leaves open the possibility of invariance, suitably defined with respect to irrelevant information, as such a criterion. This possibility depends, however, on finding a sufficiently robust delineation of informational relevance. This is a formidable problem, which to my knowledge has not been solved as of now. Recall Kahneman and Tversky's characterization, cited above: "two versions of a choice problem that are recognized to be equivalent when shown together

should elicit the same preference even when shown separately." (Kahneman and Tversky 1984:343). Recognized by whom? By the experimenter? By the decision maker herself? And under what conditions? Whether invariance will be a suitable rationality criterion will depend a lot on how these questions are answered. As Bacharach reminds us, this is a metatheoretical question that cannot be answered within a theory of rational decision making:

> whether there is a violation of [extensionality] (and so of rationality) depends on how we, the theorist, 'cut up the world'. . .. The criterion [extensionality] can only be applied after resolving a question about what it is rational to care about. (Bacharach 2001:3)

Various attempts at answering these questions have been provided, yet none has so far won general acceptance. Sen (1986: Chap. 2) introduced the idea of an *isoinformation set* containing objects of choice taken to be similar in terms of relevant information and which will be consequently treated in the same way in actual choices and judgements. Similarity in terms of relevant information here is an intersubjectively defined notion, for which it is difficult to give clear criteria. Broome (1991) discusses invariance a matter of classifying outcomes: two outcomes belong to the same class if it is irrational to have different preferences for both. Here the criterion is subjective, as it is conditional on an agent's subjective preferences. However, it isn't very useful for the present purposes (which are different from Broome's), as the invariance criterion, which is supposed to explicate rationality, would itself depend on a notion of rationality.

Sher and McKenzie (2006) recently proposed a criterion of informational relevance of different formulations as licensing different inferences:

> When there is no choice-relevant background condition C about whose probability a listener can draw inferences from the speaker's choice between frames A and B, we say that A and B are "information equivalent". Otherwise, we say that there has been information leakage from the speaker's choice of frame, and that the frames are therefore information non-equivalent. (Sher and McKenzie 2006:469)

Yet while one might use this criterion to ascertain whether in particular situations, a certain formulation was informationally relevant – and Sher and McKenzie indeed employ it in this way for assessing experimental situations – this criterion does not lend itself for a general assessment of informational relevance, as there is no clear specification when an agent is licenced to draw inferences from the speaker's formulation.

To conclude, the currently extant literature shows that the logical notion of extensionality cannot be a necessary rationality criterion for decision-making. A notion of invariance – suitably defined on informational irrelevance – might be, yet no clear delineation of informational irrelevance has as of yet found wide acceptance. That *some* framing effects – defined on extensionality or some available notion of invariance – are rational therefore seems a plausible conclusion; yet which specific framing effects are rational and which are not remains shrouded in the ambiguity of the underlying criterion.

7 Normative Theories That Model Framing

Normative decision theories prescribe how a rational decision should be made. Most of the standard normative decision theories, as described in the previous section, at least implicitly assume a relatively strong invariance requirement. Consequently, they preclude framing effects from the set of rational decisions: if descriptions of acts, states or outcomes are equivalent (typically understood as semantic identity of informational irrelevance) then the differences between these descriptions should have no influence on a rational decision. To the extent that defenders of such theories accept the existence of framing phenomena at all, they therefore propose a distinction between theories of actual behaviour and theories of rational decisions.

In contrast to this, others argue that limited violations of invariance are *compatible* with a normatively valid model of *bounded rationality*. That is, even if most people violate invariance some of the time, some of these violations might be less problematic than others, allowing for a normatively valid model of core rationality requirements. Such theories oppose the distinction between normatively valid and descriptively adequate theories of framing. Instead, they propose that one and the same theory can describe how people actually choose under framing effects, while maintaining that such choices are in fact rational. In this section, I discuss two kinds of such theories: first, those that expand standard expected utility approaches to include legitimate invariance violations, and second those that choose a reason-based account, showing how reasoning *processes* constitute legitimate violations of invariance.

Standard expected utility theories typically exclude framing effects as irrational. Savage (1954) and Anscombe and Aumann (1963), for example, did not explicitly distinguish different presentations of the same act, state or outcome. This is why they are typically interpreted as assuming extensionality. Savage, however, discusses the *small world problem*: that people do not form *one* decision problem for their whole life *at one moment in time*, partitioning the world into all relevant contingencies then – but rather divide this big world decision into a sequence of small world decisions, each of which concerning only a much rougher partitioning of the world into states (see Hirsch Hadorn 2016). People should follow the principle

> to cross one's bridges when one come to them [which] means to attack relatively simple problems of decision by artificially confining attention to so small a world that the [expected utility] principle can be applied here. (Savage 1954:16)

Because partitioning the future states of the world differently is an important form of framing, Savage here acknowledges the potential influence of framing on decision making. This conclusion is further supported by the fact that Savage explicitly excludes certain kinds of partitions as not suitable for his prescription how to make rational decisions. For example, act-dependent state partitions are excluded from a proper decision-problem set-up (as e.g. (Jeffrey 1963):8–10, points out). Yet by acknowledging the possibility of different partitions, Savage also raises

the possibility that such different partitions influence rational decisions in different ways. Take two different partitions, S and T, where T is a more fine-grained partition than S. If preferences over acts in T satisfy the Savage axioms, there is a probability function defined over states of T and a utility function over outcomes of T. Now can we calculate utilities and probabilities for S from those in T? Savage discusses two methods of doing so, and admits that these methods do not necessarily yield the same probability assignments on states in S (Savage 1954:89), for further discussion, see (Shafer 1986):480–484). Thus, although a partition satisfies the Savage axioms, this does not guarantee that the probabilities calculated in this partition do not change when the partitioned is refined (or reduced). This is Savage's *small world problem*. Clearly, it is a particularly striking case of framing of contingencies.

Savage sought to resolve the small world problem by reference to "the grand world", i.e. an ultimately detailed refinement. This device, as he admits himself, is somewhat "tongue-in-cheek" (Savage 1954:83): it posits an atomistic view of the world, although no justification is forthcoming. Only by using the grand world as a reference point, and insisting that that probability assignment is correct which is calculated from the grant world, can Savage solve the small world problem. Without it, framing effects remain possible within his theory. To the extent that Savage's theory is interpreted as a valid normative theory, it follows that these framing effects are rational.

In contrast to the partition dependence, Jeffrey's (1963) decision theory explicitly seeks a partition-invariance calculation of the expected utility of acts. He conceives of acts, outcomes and states as propositions, and calculates the expected value of acts as the sum of values of outcomes, weighted by the *conditional* probability of outcomes, given acts. As Joyce (1999:212) shows, this approach allows us to express the utility of any disjunction as a function of the utilities of its disjuncts. Thus, the partition of acts, states or outcomes has no influence on rational decision, and framing, understood in this sense, cannot be rational. Amongst decision theorists, this is commonly seen as an advantage:

> In Jeffrey's theory … there is guaranteed agreement between grand- and small-world representations of preferences. This guarantee is precisely what Savage could not deliver. The partition invariance of Jeffrey's theory should thus be seen as one of its main advantages over Savages' theory. (Joyce 1999:122)

Scholars who do not agree with Joyce on the advantages of Jeffrey's theory have introduced modifications to allow for invariance violations that might be pragmatically, if not semantically justified (e.g. Bourgeois-Gironde and Giraud 2009). However, these extensions typically do not themselves provide a criterion to distinguish between admissible and non-admissible invariance violations (as discussed in the previous section).

An alternative route of re-introducing framing into the normative framework is to deny that the Jeffrey's notion of partition invariance can exclude all relevant cases of framing. This would require that there are partitions of the world, which do not stand in the required relationship – one partition is not the disjunct in another

Fig. 8.1 An example of ambiguity

AU2

partition. Bacharach (2001) seems to hint at such a possibility. On the one hand, he wrote, most partitions exhibit this relationship – for example, partitions with respect to

> shape, colour and position: we can easily see a mark as a triangle, as a blue triangle, as a blue triangle on the left,. . . on the other hand. . . a person can see the marks as letters and as geometric shapes, but not at the same time . . . you can't integrate these two perception. (Bacharach 2001:6)

By integration, Bacharach means that two existing partitions – e.g. F={triangle, non-triangle} and G={blue, not blue} – are combined to a new partition, e.g. $H = F \bigoplus G$ = {blue triangle, blue non-triangle, non-blue triangle, non-blue non-triangle}. But he argues that not all sets of partitions can be thus integrated. A simple example, which he mentions in the quotation above, is depicted in Fig. 8.1:

One can either see the three marks as (Greek) letters or alternatively as geometric shapes, but one cannot see them as both at the same time. Other examples that Bacharach proposes include ambiguous images like Rubin's vase or the duck/rabbit image, as well as seeing outcomes either from an "I" or a "we" perspective (Bacharach 2001).

If not all frames can be integrated, then the question how to choose when the tension between such alternative frames cannot be resolved. This is where Bacharach's *variable frame theory* applies. It suggests that in coordination games, players should select strategies by choosing their best reply in each available frame. More specifically, there is an exogenous probability measure $V(F)$ defined on frames F. $V()$ is common knowledge. A strategy profile (s_i, s_{-i}) *is a variable frame equilibrium* if, *for each frame* F, the option expected from playing s_i is subjectively best from the perspective of F against s_{-i} as perceived in F (Bacharach 2001:8–9). The optimality judgment for s_i then depends on the expected utility of playing s_i against s_{-I} in each frame F, weighted by the probability of F, $V(F)$. This theory, amongst others, explains why "conceptual boundedness" of human agents, to the extent that it results in the sharedness of frames, positively contributes to people's ability to coordinate.

The above theories show how framing effects can be incorporated into expected-utility accounts of rational decision-making. An alternative, reason-based, account seeks to identify how reasoning processes rationally influence choice. Let me briefly address how extensions of this account lead to rationalization of framing, by describing Gold and List's (2004) *path-dependent decision-making*. Their account starts from the assumption that particular presentations of decision problems lead agents to consider relevant background propositions in a particular sequence, so that different presentations lead to different consideration sequences and hence to different decision paths. Such a model produces framing effects if

(i) different decision paths produce different choices, and (ii) different decision problem presentations lead to such different-choice producing paths.

To give an illustrative example, let's consider Kahneman and Tversky's Asian disease problem again (see Sect. 3). The first, "lives saved", presentation, may induce a decision path starting with factual and normative propositions about saving lives, including normative propositions like "It is not worth taking the risk that no one will be saved" – leading the agent to choose the certain option. In contrast, the second, "lives lost", presentation, may induce a decision path starting with factual and normative propositions about loosing lives, including normative propositions like "It is unacceptable to consign some people to death with certainty" – leading the agent to choose the uncertain option.

In cases like the Asian disease problem, agents have *dispositions* both to accept propositions like "It is not worth taking the risk that no one will be saved" as well as "It is unacceptable to consign some people to death with certainty". Yet depending on the decision path taken, only some of these dispositions get actualized and consequently influence decisions. As Gold and List point out, while the propositions that the agent is disposed to accept might be inconsistent (as they are in the Asian disease case), the propositions that the agent accepts on the specific decision path taken are not. Thus agents violating invariance need only suffer from implicit inconsistencies (i.e. inconsistencies regarding propositions that the agent is disposed to accept) while avoiding explicit inconsistencies between actually accepted propositions. Because such reason-based models propose specific reasoning processes, their validity (including their normative validity) will depend on what the actual mental mechanisms are that people make use of when dealing with framed acts, states or contingencies. As I argued in Sect. 4, however, research on mechanisms has been rather neglected with respect to framing.

8 Policy Relevance: How Should Decisions Be Framed?

The literature on framing discussed in the previous sections has inspired many policy proposals for intervening in human behaviour. Three key influences on policy must be distinguished. First, framing is used to *caution* policy interventions based on the reductive approach to policy analysis. Framing, as we saw, introduces various kinds of uncertainty into decision-making, including uncertainty about people's preferences, about the effect of changing the descriptions of a decision problem, and about the rationality or irrationality of observed choices. Consequently, considerations of framing might provide support for argumentative methods to deal with uncertainty in policy analysis.

Second, framing had been used to *justify* such interventions. The basic idea here is that the various framing phenomena show people to behave irrationally in a systematic way, and therefore need help from the policymaker. Third, framing has been used as the *instrument* by which various policies propose to intervene in

people's behaviour. The basic idea here is that framing is an important factor that influences behaviour, and that policy interventions can make use of it in order to achieve their ends.

Those who stress the justificatory role of framing generally agree that (i) framing phenomena are widespread and (ii) framing effects are results of irrational decision-making.

> . . .research by psychologists and economists over the past three decades has raised questions about the rationality of many judgments and decisions that individuals make. People . . . exhibit preference reversals . . . and make different choices depending on the framing of the problem. . .. (Sunstein and Thaler 2003:1168)
>
> So long as people are not choosing perfectly, it is at least possible that some policy could make them better off by improving their decisions.(Sunstein and Thaler 2003:1163)

That is, framing is a systematic behavioural phenomenon that is accurately described by some descriptive theory (discussed in Sect. 5). However, there is a normatively valid theory of behaviour, which excludes framing effects (as described in Sect. 7). Due to the difference between actual systematic behaviour and rationally required behaviour, policy interventions that make actual behaviour more rational might be justified (for similar arguments, see Conly 2013; Ariely 2008; Trout 2005; Camerer et al. 2003).

More specifically, framing plays an important role in the justification of *nudge policies* (Thaler and Sunstein 2008). Nudges are interventions on the context in which people make decisions with the aim of steering people's behaviour into specific directions. Proponents of nudges often argue that people do not have well-defined preferences, because they change their preferences in the light of rationally irrelevant frame changes. Because people often do not have clear preferences over options, welfare assessments should take into account different criteria than their preferences. Thus the justification of nudge interventions often supported with framing phenomena: people's preferences are variant under changing descriptions of the same choice situations.

Not everybody agrees with this argument. Critics point out, with arguments related to those reviewed in Sect. 6, that framing phenomena need not be irrational, and that the irrationality judgment is often based on an overtly narrow consistency criterion (Berg 2014; Berg and Gigerenzer 2010). Other concerns, in line with those discussed in Sect. 3, might question the prevalence of framing phenomena and consequently the need for interventions. Finally, some critics wonder whether framing effects really justify interventions on behaviour, and suggest instead that education can prepare people to deal with frames better on their own (Gigerenzer 2015).

This debate about whether framing justifies policy interventions is quite separate from the ways that framing has been proposed as a tool for policy interventions. One can well imagine that even if the justificatory project failed (but some other justification of policy interventions succeeded), that such policies might still employ framing as a means of influencing people's choices, if framing should prove to be an effective means for that purpose.

Three such instrumental uses of framing can be distinguished. First, policy interventions might exploit the effect of framing in order to make people choose an option the policy maker deems optimal.

> A physician, and perhaps a presidential advisor as well, could influence the decision made by the patient or by the President, without distorting or suppressing information, merely by the framing of outcomes and contingencies. Formulation effects can occur fortuitously, without anyone being aware of the impact of the frame on the ultimate decision. They can also be exploited deliberately to manipulate the relative attractiveness of options. (Kahneman and Tversky 1984:346)

Such exploitations of framing effects have been proposed, amongst others, by the *Nudge* program (Thaler and Sunstein 2008). Examples of nudging with frames include suggestions to apply lessons from the Asian disease case to the descriptions of medical treatment alternatives, so that patients are more likely to choose that option that the policymaker considers superior (Thaler and Sunstein 2008:36–37). Another example is recent proposal by Slovic and Västfjäll (2013) how to increase charitable giving through framing. Slovic and Västfjäll diagnose a systematic "insensitivity to mass tragedy" (94) in people's behaviour: when faced with suffering of large groups of victims, for example from genocide or natural disasters, people feel comparatively less compassion and give less aid than when confronted with individual victims. They propose a psychophysical model of *psychic numbing* that describes an inverse relationship between an affective valuation of saving a life and the number of lives at risk. They also argue that this affective valuation is the basis for most intuitive moral judgments about how much effort or how many resources to devote to saving lives. Consequently, they propose corrective interventions on these moral intuitions through framing the plight of many as a many plights of different individuals, each of who deserves compassion and support. Framing, as these two examples show, has become an important argument for nudge policies, as well as one of their chief policy intervention tools.

Note that these interventions might be motivated very differently. One possibility is that people go against their own preferences and do not choose what they judge best (perhaps even due to existing framing effects). In this case, (re-)framing as policy intervention is motivated by the goal to get people to choose what they really want. Another possibility is that people act according to their own preferences, but that the policymaker would prefer if they chose differently. In that case, (re-)framing is motivated to make people choose against their own wishes.

This ambiguity in the use of framing as an instrument of influence is present even in the everyday notion of framing. In colloquial English, the notion of framing has two rather disparate meaning. On the one hand, framing means "the action, method, or process of constructing, making, or fashioning something", or the result of this activity or process (OED). On the other hand, framing can also mean "the action or process of fabricating a charge or accusation against a person; an instance of this" (OED). The crucial difference here is that between a construction *simpliciter* and a construction *with deceptive intention*. It is therefore difficult to say something general about the moral evaluation of framing policies, but it is

obvious that at least *some* uses of framing in this way are not compatible with liberal values (Grüne-Yanoff 2012).

Another use of our knowledge of framing effects as a policy tool is to design choice environments in such a way that framing effects are neutralized or eliminated whenever possible. This requires the idea that some frames exert less strong influences on reasoning and decision than others – i.e. that there is a canonical frame. Kahneman and Tversky suggest something along these lines, when they recommend to

> adopt a procedure that will transform equivalent versions of any problem into the same canonical representation. This is the rationale for the standard admonition to students of business, that they should consider each decision problem in terms of total assets rather than in terms of gains or losses (Schlaifer 1959). Such a representation would avoid the violations of invariance illustrated in the previous problems, but the advice is easier to give than to follow. (Kahneman and Tversky 1984:344)

One possible basis for such a neutrality argument is the hypothesis that human cognition is well adapted to certain kinds of representations, but not to others. With respect to statistical inference, for example, some have argued that our cognitive algorithms are not adapted to probabilities or percentages, as these concepts and tools have been developed only rather recently. Consequently, policies should aim to design inference or choice tasks with representations that people are most adapted to. In the case of statistical inference, Gigerenzer and Hoffrage (1995) and Hoffrage et al. (2000) showed that statistics expressed as *natural frequencies* improve the statistical reasoning of experts and non-experts alike.[2] For example, advanced medical students asked to solve medical diagnostic tasks performed much better when the statistics were presented as natural frequencies than as probabilities. Similar results have been reported for medical doctors (in a range of specialties), HIV counsellors, lawyers, and law students (Anderson et al. 2012; Akl et al. 2011; Lindsey et al. 2003; Hoffrage et al. 2000).

Bacharach seems to consider a similar idea when he suggests that many frames might be integrable: by providing a finer partition, two seemingly conflicting perspectives on the world can be combined in a more detail-rich frame. However, it remains unclear why this frame should be considered more 'neutral' than either of the original ones. What remains true is that "one does not just see, but one sees as" (Bacharach 2001:1); hence the neutral frame might remain a chimera.

A third use of our knowledge of framing effects as a policy tool – particularly if the first one is ethically questionable and the second one unachievable – is to elicit reflection through reframing. That is, the policy maker might present decision makers who are prone to framing effects with relevant information in *different* formats at the same time. In effect, this seeks to test the robustness of preferences by

[2] Natural frequencies refer to the outcomes of natural sampling—that is, the acquisition of information by updating event frequencies without artificially fixing the marginal frequencies. Unlike probabilities and relative frequencies, natural frequencies are raw observations that have not been normalized with respect to the base rates of the event in question.

deliberate attempts to frame a decision problem in more than one way (cf. Fischhoff et al. 1980). Such an approach instead of nudging or neutralising, seeks to *boost* people's abilities to deal with informationally and representationally challenging situations (Grüne-Yanoff and Hertwig 2015). The boost approach aims to enhance people's ability to understand and see through confusing and misleading representations by making those representations less manipulative and opaque, rendering them less computationally demanding (Gigerenzer and Hoffrage 1995), and making them semantically and pragmatically less ambiguous (Hertwig and Gigerenzer 1999). From the boost perspective, difficulties understanding statistical information are seen not as an incorrigible mental deficiency of, say, doctors or patients, but as largely attributable to poor or intentionally misleading information. Moreover, the goal is not to push people toward a particular goal (e.g., to seek or not seek a particular treatment), but to help everybody (e.g., doctors and patients) to understand statistical information as the first critical step toward figuring out one's preference.

9 Conclusion

Framing is an important set of phenomena that challenges the standard theories of rational decision making and the notions of rationality they propose. Because framing seemingly drives a wedge between actual behaviour and normative standards imposed on behaviour, it has been used as a justification for policies intervening in behaviour. Nevertheless, many questions remain. From the survey of experimental elicitation, it isn't obvious how unified the notion of framing is, nor is it obvious that it is as prevalent as sometimes claimed. From the survey of mechanistic models and descriptive theories it appears that many questions when and how framing effects behaviour are not fully settled. Furthermore, there is considerable controversy to what extent the sensitivity of decisions to framing is irrational. Finally, consideration of framing might provide support for argumentative methods in policy analysis. All these questions have import on whether policies intervening on framing are justifiable, as well as whether framing is an effective and morally permissible tool of policy making.

Recommended Readings

Arrow, K. J. (1982). Risk perception in psychology and economics. *Economic Enquiry*, *20*, 1–9.

Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*, 275–305.

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* (New Series), *211*, 453–458.

References

Ahn, D., & Ergin, H. (2010). Framing contingencies. *Econometrica, 78*, 655–695.

Akl, E. A., Oxman, A. D., Herrin, J., Vist, G. E., Terrenato, J., Sperati, F., Costiniuk, C., Blank, D., & Schünemann, H. (2011). Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database of Systematic Reviews*. doi:10.1002/14651858.CD006776.pub2.

Anderson, B. L., Gigerenzer, G., Parker, S., & Schulkin, J. (2012). Statistical literacy in obstetricians and gynecologists. *Journal for Healthcare Quality, 36*, 5–17.

Anscombe, F. J., & Aumann, R. J. (1963). A definition of subjective probability. *Annals of Mathematical Statistics, 34*, 199–205.

Ariely, D. (2008). *Predictably irrational: The hidden forces that shape our decisions* (1st ed.). New York: HarperCollins.

Arrow, K. J. (1982). Risk perception in psychology and economics. *Economic Enquiry*, *20*, 1–9.

Bacharach, M. O. (2001). Framing and cognition in economics: The bad news and the good. Lecture notes for the *ISER Workshop, Cognitive Processes in Economics*. http://cess-wb.nuff.ox.ac.uk/documents/mb/lecnotes.pdf.

Bacharach, M., Gold, N., & Sugden, R. (2006). *Beyond individual choice: Teams and frames in game theory*. Princeton: Princeton University Press.

Berg, N. (2014). The consistency and ecological rationality approaches to normative bounded rationality. *Journal of Economic Methodology, 21*, 375–395.

Berg, N., & Gigerenzer, G. (2010). As-if behavioral economics: Neoclassical economics in disguise? *History of Economic Ideas, 18*, 133–166.

Berwick, D. M., Fineberg, H. V., & Weinstein, M. C. (1981). When doctors meet numbers. *American Journal of Medicine, 71*, 991–998.

Bourgeois-Gironde, S., & Giraud, R. (2009). Framing effects as violations of extensionality. *Theory and Decision, 67*, 385–404.

Broome, J. (1991). *Weighing goods: Equality, uncertainty and time*. Oxford: Wiley-Blackwell.

Brun, G., & Betz, G. (2016). Analysing practical argumentation. In S. O. Hansson & G. Hirsch Hadorn (Eds.), *The argumentative turn in policy analysis. Reasoning about uncertainty* (p. xx). Cham: Springer.

Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Regulation for conservatives: Behavioral economics and the case for "Asymmetric Paternalism". *University of Pennsylvania Law Review, 151*, 1211–1254.

Conly, S. (2013). *Against autonomy: Justifying coercive paternalism*. Cambridge: Cambridge University Press.

Doorn, N. (2016). Reasoning about uncertainty in flood risk governance. In S. O. Hansson & G. Hirsch Hadorn (Eds.), *The argumentative turn in policy analysis. Reasoning about uncertainty* (p. xx). Cham: Springer.

Edvardsson Björnberg, K. (2016). Setting and revising goals. In S. O. Hansson & G. Hirsch Hadorn (Eds.), *The argumentative turn in policy analysis. Reasoning about uncertainty* (p. xx). Cham: Springer.

Elliott, K. C. (2016). Climate geoengineering. In S. O. Hansson & G. Hirsch Hadorn (Eds.), *The argumentative turn in policy analysis. Reasoning about uncertainty* (p. xx). Cham: Springer.

Evans, G. W., & Crumbaugh, C. M. (1966). Effects of prisoner's dilemma format on cooperative behavior. *Journal of Personality and Social Psychology, 3*, 486.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1980). Knowing what you want: Measuring labile values. In T. Wallsten (Ed.), *Cognitive processes in choice and decision behavior* (pp. 117–141). Hillsdale: Erlbaum.

Fox, C. R., & Rottenstreich, Y. (2003). Partition priming in judgment under uncertainty. *Psychological Science, 14*, 195–200.

Gallagher, K. M., & Updegraff, J. A. (2012). Health message framing effects on attitudes, intentions, and behavior: A meta-analytic review. *Annals of Behavioral Medicine, 43*, 101–116.
Gambara, H., & Piñon, A. (2005). A meta-analytic review of framing effect: Risky, attribute and goal framing. *Psicothema, 17*, 325–331.
Gigerenzer, G. (2015). On the supposed evidence for libertarian paternalism. *Review of Philosophy and Psychology, 6*, 361–383.
Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science, 1*, 107–143.
Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684–704.
Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Cambridge, MA: Harvard University Press. AU4
Gold, N., & List, C. (2004). Framing as path-dependence. *Economics and Philosophy, 20*, 253–277.
Grüne-Yanoff, T. (2012). Old wine in new casks: Libertarian paternalism still violates liberal principles. *Social Choice and Welfare, 38*, 635–645.
Grüne-Yanoff, T. (2015). Why behavioural policy needs mechanistic evidence. *Economics and Philosophy*, (in press).
Grüne-Yanoff, T., & Hertwig, R. (2015). Nudge versus boost: How coherent are policy and theory? *Minds and Machines*. doi:10.1007/s11023-015-9367-9.
Grunwald, A. (2016). Synthetic biology: Seeking for orientation in the absence of valid prospective knowledge and of common values. In S. O. Hansson & G. Hirsch Hadorn (Eds.), *The argumentative turn in policy analysis. Reasoning about uncertainty* (p. xx). Cham: Springer.
Hansson, S. O., & Hirsch Hadorn, G. (2016). Introducing the argumentative turn in policy analysis. In S. O. Hansson & G. Hirsch Hadorn (Eds.), *The argumentative turn in policy analysis. Reasoning about uncertainty* (p. xx). Cham: Springer.
Hershey, J. C., & Schoemaker, P. J. H. (1980). Risk taking and problem context in the domain of losses: An expected-utility analysis. *Journal of Risk and Insurance, 47*, 111–132.
Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*, 275–305.
Heukelom, F. (2014). *Behavioral economics: A history*. Cambridge: Cambridge University Press.
Hirsch Hadorn, G. (2016). Temporal strategies for decision making. In S. O. Hansson & G. Hirsch Hadorn (Eds.), *The argumentative turn in policy analysis. Reasoning about uncertainty* (p. xx). Cham: Springer.
Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science, 290*, 2261–2262.
Jeffrey, R. C. (1963). *The logic of decision*. Chicago: University of Chicago Press.
Joyce, J. M. (1999). *The foundations of causal decision theory*. Cambridge: Cambridge University Press.
Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430–454.
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.
Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*, 341–350.
Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative and methodological challenges. *Behavioral and Brain Sciences, 19*, 1–53.
Kühberger, A. (1995). The framing of decisions: A new look at old problems. *Organizational Behavior and Human Decision Processes, 62*, 230–240.
Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes, 75*, 23–55.

Le Menestrel, M., & Wassenhove, L. V. (2001). The domain and interpretation of utility functions: An exploration. *Theory and Decision, 51*, 329–349.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes, 76*, 149–188.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89*, 46.

Lindsey, S., Hertwig, R., & Gigerenzer, G. (2003). Communicating statistical DNA evidence. *Jurimetrics: The Journal of Law, Science, and Technology, 43*, 147–163.

Mandel, D. R. (2001). Gain-loss framing and choice: Separating outcome formulations from descriptor formulations. *Organizational Behavior and Human Decision Processes, 85*, 56–76.

Möller, N. (2016). Value uncertainty. In S. O. Hansson & G. Hirsch Hadorn (Eds.), *The argumentative turn in policy analysis. Reasoning about uncertainty* (p. xx). Cham: Springer.

Oxford English Dictionary (OED). Sept 2011. "framing, n.". Oxford University Press. http://dictionary.oed.com/. Accessed 30 Sept 2014.

Pratt, J. W., Wise, D., & Zeckhauser, R. (1979). Price differences in almost competitive markets. *Quarterly Journal of Economics, 93*, 189–211.

Rubinstein, A. (2000). *Modeling bounded rationality*. Cambridge: MIT Press.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Sen, A. (1986). Information and invariance in normative choice. In W. P. Heller, R. M. Starr, & D. A. Starret (Eds.), *Social choice and public decision making* (Essays in Honor of Kenneth J. Arrow, Vol. 1, pp. 29–55). Cambridge: Cambridge University Press.

Shafer, G. (1986). Savage revisited. *Statistical Science, 1*, 463–485.

Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition*, *101*, 467–494.

Slovic, P., & Västfjäll, D. (2013). The more who die, the less we care: Psychic numbing and genocide. In A. Olivier (Ed.), *Behavioural public policy* (pp. 94–109). Cambridge: Cambridge University Press.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1982). Response mode, framing, and information-processing effects in risk assessment. In R. Hogarth (Ed.), *New directions for methodology of social and behavioral science: Question framing and response consistency* (pp. 21–36). San Francisco: Jossey-Bass.

Sunstein, C. R., & Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, 1159–1202.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization, 1*, 39–60.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge*. New Haven: Yale University Press.

Trout, J. D. (2005). Paternalism and cognitive bias. *Law and Philosophy, 24*, 393–434.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science* (New Series), *211*, 453–458.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business, 59*, S251–S278.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101*, 547–567.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.