# CHAPTER 19

# Judgment Heuristics

## 19.1 Introduction

The modern *heuristics and biases* research program commenced with the seminal articles by Tversky and Kahneman (1971, 1974). A central insight of this program is that individuals are not fully rational Bayesian decision makers, but their behavior is not random either. Individuals often ignore a substantial part of the available information or use it selectively to implement simple *rules of thumb* or *heuristics*. Most heuristics are *fast*, in terms of computation time required, and *frugal* in the use of information. These heuristics do not optimize in the classical sense, hence, their performance is not necessarily optimal. However, when there is uncertainty about the optimal solution, their performance can be at least as good or even better than some more complex methods (Gigerenzer et al., 1999).

To motivate their approach, Tversky and Kahneman (1974) give the following example. Individuals often judge distance by the heuristic of how clear it is; sharper objects are thought to be closer. This heuristic is fast and frugal. While this heuristic often works well, it might also sometimes give rise to misleading perceptions. For instance, when atmospheric conditions are hazy, distances might be overestimated. Conversely, when conditions are very clear, distances might be underestimated; for a revised account of this example, see Kahneman and Frederick (2005).

The heuristics and biases approach often considers problems that require some probabilistic assessments or inferences. Classical economics assumes that individuals respect the classical laws of probability and statistical inference. In contrast, the evidence indicates that in these situations, individuals employ a range of *judgment heuristics*. The heuristics and biases approach demonstrates a systematic departure of the actual behavior of individuals from the predictions of the laws of classical statistics.[1] It is notable that even statistically sophisticated researchers and experts typically rely on these judgment heuristics.[2] The heuristics and biases approach is quite distinct from the formal decision theory approach in Part 1 of the book. It does not necessarily rely on any explicit optimization method, although most formal applications use a mixture of optimization methods in conjunction with simple rules of thumb.

We begin with the *representativeness heuristic* in Section 19.2 (Tversky and Kahneman, 1974). When individuals use this heuristic, they compare the likeness of a sample, even if the sample is

[1] There are many excellent surveys and collected readings. Particularly relevant to economists are Camerer (1995), Gilovich et al. (2002), Kahneman et al. (1982), Kahneman (2003), and Kahneman (2011). For a commendable effort at using some of this machinery to analyze development issues, see WDR (2015).

[2] See for instance, Gilovich et al. (2002), Kahneman (2003), Kahneman (2011), and Tetlock (2006).

small, with known features of the population distribution. The evidence shows that relative to a Bayesian, users of the representativeness heuristic assign too high a probability that small samples share the properties of large samples (Camerer 1987, 1990, 1995). Such individuals behave as if they obeyed a *law of small numbers*, rather than the statistically correct, *law of large numbers*.

Reliance on the law of small numbers has several implications. Individuals find it difficult to produce sequences of outcomes of random variables, and their constructed sequences show up too much *negative autocorrelation*, which is also sometimes known as *alternation bias* or *negative recency* (Tversky and Kahneman, 1974; Bar-Hillel and Wagenaar, 1991; Rapoport and Budescu, 1997). A belief in negative autocorrelation, when there is none in the underlying distribution, in betting domains is known as the *gambler's fallacy*. For instance, gamblers are often reluctant to bet on the immediately previous winning numbers in a lottery, yet the conditional and unconditional mathematical probabilities of these numbers is unchanged. Evidence from betting behavior gives strong support to the gambler's fallacy (Metzger, 1984; Clotfelter and Cook, 1993; Terrell, 1994; Croson and Sundali, 2005), as does the evidence from the behavior of investors in stock markets (Odean, 1998).

The converse of the gambler's fallacy, i.e., a belief in *positive autocorrelation* in outcomes, when there is none, is known as the *hot hands fallacy*. When subjects observe too many consecutive outcomes of one kind, they come to believe that the realization of a similar outcome is more probable. For instance, basketball players who score heavily in a game are often perceived to be on a hot streak even when there is no evidence of underlying hot hands (Gilovich et al., 1985; Tversky and Gilovich, 1989a,b; Camerer, 1989). Similar beliefs are often assigned to winning football teams. Recent evidence indicates that for basketball and baseball, there might be an underlying hot hands phenomenon (Green and Zwiebel, 2013; Bocskocsky et al., 2014; Miller and Sanjurjo, 2015), yet we argue that it does not invalidate the hot hands fallacy.

In contrast to outcomes from sports, there is good support for the hot hands fallacy when the underlying stochastic process is inanimate. For instance, lotto stores that sell winning lotto tickets experience unusually high sales for up to 40 weeks following the lotto win (Guryan and Kearney, 2008). Individuals are willing to pay more for experts who are believed to possess hot hands (Powdthavee and Riyanto, 2015). Players who win in casinos are less likely to quit, believing that they are on a winning streak (Croson and Sundali, 2005). When the underlying stochastic process relies on human skills, one is more likely to observe a gambler's fallacy and when the process is inanimate, a hot hands fallacy is more likely to be observed (Ayton and Fischer, 2004).

The gambler's fallacy and the hot hands fallacy are both likely to be caused by the representativeness heuristic. An individual who initially believes the underlying process to be random (so zero autocorrelation in successive draws) predicts too much negative autocorrelation to try to make outcomes appear random (gambler's fallacy). However, when too many outcomes of the same type are observed, the individual starts to believe that the underlying process is positively autocorrelated, even when small samples do not statistically justify this conclusion (hot hands fallacy).

Finally, we give a useful formalization of the law of small numbers due to Rabin (2002). This model also explains the existence of *overconfidence* and the gambler's fallacy.

Section 19.3 considers the *conjunction fallacy*, namely, the tendency to assign higher probabilities to the intersection of two sets relative to any of the two sets alone. The most well known and critically discussed illustration of this phenomenon is the *Linda problem* (Tversky and Kahneman, 1983). After describing a set of features of a hypothetical female called Linda (e.g., concerned with social justice, participated in anti-nuclear demonstrations) subjects in experiments are asked to judge Linda's profession. A large number of subjects chose the composite description

"Linda is a bank teller and active in the feminist movement" relative to either of the two descriptions separately. Since the probability of the intersection cannot ever be larger than any of the individual sets, this is termed as the conjunction fallacy.

Hertwig and Gigerenzer (1999), two long-standing critics of the Kahneman–Tversky approach, claimed that the conjunction fallacy disappears if it is posed in terms of natural frequencies. However, in a between-subjects design, when the Linda problem was presented in a natural frequency format, the conjunction fallacy survived (Kahneman and Tversky, 1996). The conjunction fallacy was shown to be reduced in the presence of incentives and joint decisions made by groups of two and three individuals (Charness et al. 2010). However, other experimental evidence shows that the conjunction fallacy survives intact not only in the presence of monetary incentives but also when the same problem is framed in alternative ways (Bonini et al., 2004; Stolarz-Fantino et al., 2003). Another potential explanation of the conjunction fallacy is given in terms of the potential confound between the word "and" and the logical operator $\wedge$ (Camerer, 1995; Hertwig et al., 2008). However, other research shows that while the conjunction fallacy may be lessened slightly due to these reasons, it cannot be eliminated (Tentori et al., 2004; Tentori and Crupi, 2012).

Section 19.4 considers the *availability heuristic*. Individuals judge the probability of an event by the ease with which similar events are accessible or "available" to the mind (Tversky and Kahneman, 1974). Availability may depend on how salient the event is, or how vivid or lasting an impression it made, or even how easy it is to construct hypothetical scenarios. Media coverage is one important determinant of the availability of events (Lichtenstein et al., 1978; Eisenman, 1993). An influential early empirical study established a firm link between the perceived frequency of deaths by various causes and proxies for the availability of such events in memory (Lichtenstein et al., 1978). These results have been replicated and extended (Hertwig et al., 2005; Pachur et al., 2012). Indeed, there is a close relation between the availability heuristic and the *affect heuristic* that we consider later (Pachur et al., 2012).

Section 19.5 considers the *affect heuristic* and highlights the link between emotions and heuristics (Zajonc, 1980; Slovic et al., 2002; Loewenstein et al., 2001). The basic idea is that we classify and tag events and memories by their *affect*, for instance, positive affect (enjoyable holiday) or negative affect (painful skiing accident). In making judgments and decisions, we are then influenced by the pool of these positive and negative affects or the *affect-pool*. For instance, in buying a car, we might focus a great deal on the affect that the car has on us, for instance, a sleek car, a cool car, or an enviable car. The affect heuristic substitutes a hard question (what do I think about it?) by an easier question (what do I feel about it?) that allows for quicker decision making (Kahneman, 2011).

Evidence shows that our response to various hazards, such as smoking and nuclear power, depends on the affect, or *dread* in this case, that they invoke (Fischhoff et al., 1978; Finucane et al., 2000). Experts also exhibit the affect heuristic; the perceived riskiness of various hazardous substances is closely related to the affect that they have on experts (Slovic et al., 1999). Presenting the data in terms of natural frequencies to experts enhances the affect of an event (Slovic et al., 2000).

*Anchoring*, considered in Section 19.6, is one of the most robust and important heuristics that people use. Anchoring occurs when people are influenced in their judgments and decisions by an initial suggestion, or *anchor*, that may be irrelevant to the question asked. Anchors may be self-generated or provided by the experimenter. Anchoring has been observed in a wide variety of phenomena. Furthermore, once individuals fix an anchor in their minds, they adjust too slowly, and insufficiently, towards the target; this is particularly the case with self-generated anchors (Epley and Gilovich, 2001). However, when the anchors are provided by the experimenter, then

individuals typically engage System 1 of the brain (see Section 19.11) in gauging the accuracy of the anchor. System 1 is biased towards finding confirmatory evidence, which leads to anchoring.

Early studies on anchoring considered the relation between stimulus and sensation. However, Tversky and Kahneman (1974) revolutionized the field by focusing on numeric anchors, which is now the established method of eliciting anchoring effects. In a famous experiment, they rigged a wheel of fortune with numbers 1–100 to come up with either 10 or 65 (low and high anchors). They first asked subjects if the number of African nations in the United Nations was larger or smaller than the anchor (comparative judgment task). Then they asked subjects to guess the number of African nations in the United Nations (absolute judgment task). Although, in this case, the anchors are irrelevant to the questions, they had a significant effect on the subjects' guesses.

Experienced judges are also subject to the anchoring effect (Englich and Mussweiler, 2001). Anchors that are clearly implausible also have a significant effect on subjects' choices (Strack and Mussweiler, 1997; Mussweiler and Strack, 2000a). Field data show that estate agents are unduly influenced by the initially suggested list price for a house (Northcraft and Neale, 1987). Caps on damages awarded to litigants serve as an anchor and have an effect on the rate of pretrial settlements (Babcock and Pogarsky, 1999) as well as on the level of damages (Pogarsky and Babcock, 2001).

Anchoring is one of the most robust heuristics and has been documented in a wide range of contexts such as poetry reading sessions (Ariely et al., 2006); violation of a basic implication of efficient markets (Mussweiler and Schneller, 2003); and first offers in price negotiation (Galinsky and Mussweiler, 2001). Anchoring has also been used to explain a range of other behavioral phenomena such as hindsight bias, non-linear probability weighting, and preference reversals.

Section 19.7 shows that the behavior of individuals is inconsistent with Bayes' rule, particularly when they face probabilistic information. In particular, individuals engage in *base rate neglect*, which is the tendency to give insufficient attention to the prevalence of a particular trait or feature in the overall population. Consider the well-known cab problem (Kahneman and Tversky, 1972; Bar-Hillel, 1980; Tversky and Kahneman, 1980). Cabs in a city are either green (85%) or blue (15%). One of the cabs was involved in an accident last night. A witness with a reliability rate of 80% came forward to identify that the cab involved in the accident was blue. What is the probability that a blue cab was involved? The correct answer, using Bayes' rule, is 41.4%, while the modal experimental response was 80%. Subjects appear to ignore the fact that only 15% of the cabs are blue (the base rate); thus, they engage in base rate neglect.

The base rate underweighting phenomenon is robust to several perturbations, although some might lessen it. For instance, base rate underweighting is found even when base rates are made more salient (Ajzen, 1977; Bar-Hillel, 1980), or when subjects are highly educated (Casscells et al., 1978), or when individuals have higher measured intelligence (Stanovich and West, 2000). A much better predictor of base rate underweighting is the important distinction between *causal* and *incidental* base rates (Ajzen, 1977). An example of a causal base rate is the instruction: "there are an equal number of Blue and Green taxis but 85% of the taxis involved in accidents are Green." In this case, base rate neglect is reduced (Kahneman, 2011). Furthermore, when information on base rates is combined with other non-causal information, then base rate neglect may also be reduced (Bar-Hillel, 1980).

As with many other heuristics, the claim is made that presenting the information in a *frequency format* rather than a *probability format* reduces base rate neglect (Brase, 2002b; Cosmides and Tooby, 1996; Gigerenzer and Hoffrage, 1995; Pinker, 1997). For most empirical studies that use the frequency format, the percentage of subjects who do not engage in base rate underweighting is about 40% (Barbey and Sloman, 2007). While this is an improvement over the probability

format, the case for the efficacy of frequency format is overstated. Even when data are given in natural frequencies, but the set inclusion relation is not made obvious, the benefits of natural frequency representation drop significantly (Girotto and Gonzalez, 2001). An even more practical consideration in the debate over natural frequencies versus the probability format, based on anecdotal evidence, is that most real-world economic information of interest is arguably presented in a probability format.

In contrast to base rate underweighting, the phenomenon of *conservatism* refers to *underweighting of the likelihood of a sample* (Camerer, 1995). In Section 19.7.3, we offer a reconciliation of base rate neglect with conservatism by distinguishing between the *strength* and the *weight* of *evidence* (Griffin and Tversky, 1992). For instance, in a reference letter, strength refers to the warmth and support for the applicant, while weight refers to the standing and credibility of the referee. Individuals typically give relatively greater prominence to strength. This is a form of overconfidence or base rate neglect. However, when the strength of the letter is low relative to the weight one might feel underconfident, which is a form of conservatism.

In Section 19.8, we consider *hindsight bias*. This occurs when individuals report, ex-post, that actual events coincided with their initial predictions of the events, i.e., *they knew it all along*. Hindsight bias occurs in a between-subjects (Fischhoff, 1975) and a within-subjects design (Fischhoff and Beyth, 1975). Hindsight bias differs from learning. Under learning, one learns to make better judgments about the future. Under hindsight bias, one misperceives one's own predictions made in the past, in a manner that gives rise to a lower predictive error. The main explanation of hindsight bias is that the human mind has only partial ability to recall what it used to believe in the past, particularly when one changes one's mind (Kahneman, 2011). Hindsight bias might also arise as a coping mechanism to make sense of an unpredictable world, or to signal one's competence to others (Rachlinski, 1998).

Hindsight bias is particularly important in legal judgments because jurors and judges need to determine the culpability of a defendant's action when they already know it has caused harm (Rachlinski, 1998). Hindsight bias has been found in very diverse contexts. A meta-study shows that experts are subjected to the hindsight bias (Guilbault et al., 2004). In Section 19.8.4, we apply hindsight bias to a problem in finance (Biais and Weber, 2009). We show that under hindsight bias, decision makers reduce estimates of volatility and engage in less corrective activity of incorrect models. Empirical evidence bears out the predictions of the model.

Section 19.9 considers *confirmation bias* in which individuals engage in selective interpretation of evidence so that it supports their initial beliefs. This is sometimes known as *biased assimilation* in psychology (Nickerson, 1998). There is extensive evidence of confirmation bias (Lord et al., 1979; Darley and Groos, 1983; Munro et al., 2002). Confirmation bias is not mitigated by greater cognitive abilities (Stanovich and West, 2007, 2008a,b). We also sketch a formal model of confirmation bias that can also explain the prevalence of overconfidence and underconfidence (Rabin and Schrag, 1999).

Section 19.10 considers a range of other judgment heuristics that include the following.[3] People often do not take account of the phenomenon of *regression to the mean* (Galton, 1886; Tversky and

[3] It is clearly not possible to consider all possible heuristics. For instance, we omit a discussion of the heuristics used to form macroeconomic expectations. Rational expectations still lie at the heart of modern macroeconomic theory and macroeconomic forecasting. Fuster et al. (2010) review evidence that is not consistent with rational expectations and outline an alternative that they call *natural expectations*. Beshears et al. (2013b) show that individuals often find it quite hard to recognize the dynamics of mean-reverting processes, an ability that is essential for existing macroeconomic models of forecasting. Ericson et al. (2015) propose a heuristic approach to choosing among time–outcome pairs that outperforms exponential and hyperbolic discounting. These approaches are likely to gain increasing importance in the future, possibly with profound effects on macroeconomics.

Kahneman, 1974). If outcomes are influenced by luck, then better (worse) than average outcomes in the current period should be expected to revert to more modest (better) levels in the future. For instance, on average, golfers with high (low) initial scores on day 1 of a tournament report lower (higher) scores on day 2 (Kahneman, 2011). People are often not very good at distinguishing between *necessary and sufficient conditions*, a distinction that is critical in many economic models (Wason, 1968). *Attribute substitution* occurs when one attribute of an object is substituted by another (Kahneman and Frederick, 2005). For instance, the answers to the two questions: "How happy are you with your life in general?" and "How many dates did you have last month?" depend on the order in which the questions are asked (Strack et al., 1988).

Section 19.11 focuses on one potential explanation of judgment heuristics. It is useful to think of the human brain as made up of two broad, but purely hypothetical, systems; System 1 and System 2 (Stanovich and West, 2000, 2002). System 1, or the emotional system, is fast, automatic, reactive, and requires little mental effort; it is also never turned off. System 2, or the cognitive system, is slow, deliberative, uses scarce cognitive resources, and takes account of the long-run consequences of our actions. Kahneman (2011) offers a powerful and cogent case for explaining judgment heuristics using this two systems approach.

Since System 1 is automatic, most judgment heuristics arise from its response. Even when an unusual situation is encountered and the services of System 2 are called upon, the agenda is often set by System 1, which controls the events that are recalled. System 1 calls the shots most of the time. This is revealed by the biases exhibited by trained statisticians in dealing with questions on probability (Tversky and Kahneman, 1973), and the extensive evidence from other expert judgments that is summarized in Section 19.18. System 1 is relatively more likely to be engaged when decisions have to be made quickly, statistical training is low, and cognitive ability is lower (Kahneman and Frederick, 2005; Oechssler et al., 2009).

In Section 19.12, we use the model of *thinking and persuasion* as a possible formalization of judgment heuristics (Shleifer et al., 2008). The idea is that, on account of cognitive limitations, individuals group different events into categories (Mullainathan, 2002). They then treat all events in the same category in an identical manner, so that there is across-category heterogeneity and within-category homogeneity; such thinkers may be termed as *coarse thinkers*. We consider the possibility that there is a persuader who controls the message received by a coarse thinker. It is never optimal for a persuader to send a misleading message to a classical Bayesian thinker but it might be optimal to mislead a coarse thinker. This model can address several puzzling phenomena such as advertising and different forms of persuasion by politicians, the media, firms, and experts. Empirical evidence supports the existence of persuasions from many domains (DellaVigna and Gentzkow, 2010).

In Section 19.13, we consider *mental models* that people form about the world around them and their place in it. These models can take several forms such as beliefs about causal relations, social identities, categorizing disparate information into coarse categories, and social worldviews (Denzau and North, 1994; WDR, 2015). Mental models are pervasive, help economize on cognitive costs, are transmitted from generation to generation, once formed they are typically inertial, and can both assist and deter optimal decisions, depending on their accuracy (Datta and Mullainathan, 2014). Humans are hardwired with some mental models, while in other cases, history plays an important role in their formation (Nunn, 2008; Alesina et al., 2013; Acemoglu et al., 2013). We also consider attempts to transform incorrect mental models into more useful ones (Datta and Mullainathan, 2014).

Section 19.14 discusses Herbert Simon's seminal approach to bounded rationality (Simon, 1978). The idea is that in the presence of cognitive limitations, individuals should not be expected

to engage in the optimization exercises that neoclassical economic theory assumes to be the default (*substantive rationality*). As Simon (1978) famously noted: "But there are no direct observations that individuals or firms do actually equate marginal costs and revenues." Thus, he recommends concentrating on *procedural rationality*, which focuses on the quality of the process of decisions. As Reinhard Selten has argued on many occasions, the problem is that economic problems are typically hard in the sense that they are *NP-Complete*, a term that is borrowed from the computer science literature.[4] In an NP-Complete problem, the number of steps needed to solve a problem in any algorithm of the solution, grows exponentially with the size of the problem (e.g., the travelling salesman problem[5]). So, even for moderately difficult problems of this sort, the computing time required to solve the problem is very high. It then becomes a leap of faith that Homo economicus can solve these problems or act as if he/she can.

An example of a process that satisfies procedural rationality is *satisficing behavior*, where individuals have an *aspiration level* and adjust gradually in its direction through a sequence of steps (Simon, 1955, 1978, 2000; Selten, 2001). The word "satisficing" is a neologism that alludes to the fact that such decision procedures are satisfactory and they suffice. Empirical evidence is supportive of the theory (Caplin et al., 2011); subjects appear to search sequentially and stop searching when their aspiration level is achieved.

We demonstrate how the cooperative outcome can be achieved in the prisoner's dilemma game as one's aspirations adjust over time (Karandikar et al., 1998; Cho and Matsui, 2005; Börgers and Sarin, 2000). To understand the basic idea, suppose that in the prisoner's dilemma game, players have an aspiration level that lies in between the payoffs from pure defection and pure cooperation. Then, whenever one or both players defect, one of the players finds that the payoff is below the aspiration level. Such a player then experiments by choosing another strategy. These experiments eventually lead to the cooperative outcome where all experimentation ceases.

Section 19.14.2 considers the approach of Gerd Gigerenzer and colleagues on fast and frugal heuristics. The focus of this work is to try to establish the following (Gigerenzer et al., 1999; Gigerenzer and Selten 2001): Heuristics are not mistakes, but a rational response to bounded rationality, and heuristics often outperform selected optimization methods. In some cases these heuristics are derived from evolutionary adaptation. For instance, individuals use the *gaze heuristic*, to catch a ball by employing a constant angle between the ball and the gaze. However, there should be no presumption that evolution has prepared us with the relevant heuristics needed to optimally invest in alternative pension plans or choose among alternative stocks.

We consider a range of fast and frugal heuristics, but mainly the *recognition heuristic*, and the *take-the-best* heuristic. Using the recognition heuristic, between two stocks, stock market participants invest in the stock that they recognize, relative to the one that they do not. There is some evidence that this method outperforms several investment strategies, such as purely random selection of stocks, expert judgment, and a selection of statistical selection techniques (Borges et al., 1999; Czerlinski et al., 1999). The recognition heuristic has also been used to determine which city has a higher population (Gigerenzer and Goldstein, 1996). However, doubts have

[4] See, for instance, a video of the entertaining panel discussion on behavioral economics at the 2011 Lindau Nobel Laureate Meeting in Economic Sciences. The participants were George A. Akerlof, Robert J. Aumann, Eric S. Maskin, Daniel L. McFadden, Edmund S. Phelps, and Reinhard Selten.

[5] Suppose that a salesman is provided with a list of cities and the distances between each pair of cities. The objective, starting with a home city, is to find the shortest possible route such that the salesman visits each city exactly once and returns to the home city.

also been raised about the robustness of these results, and contrary evidence found (Oppenheimer, 2003).

There are several concerns about this literature. First, it demonstrates the superiority of fast and frugal heuristics in domains where the relevant optimization benchmark is typically unclear. For instance, in predicting stock market prices, it is well known that the efficient markets hypothesis fails, so what is the relevant optimal solution to compare with the outcome arising from the heuristics? Furthermore, these heuristics are experimenter-provided. No compelling evidence is provided whether people do actually use these heuristics. Furthermore, in many complicated economic problems, it is not clear what these heuristics should be. In contrast, the Kahneman–Tversky approach addresses squarely the issue of which heuristics are used in different environments and the relevant optimization benchmark is unambiguous.

Section 19.15 considers what has come to be known as the *great rationality debate* in psychology, whose main protagonists are viewed as Kahneman–Tversky on the one hand, and Gerd Gigerenzer/colleagues on the other (Stanovich and West, 2000; Stanovich, 2012). The Kahneman–Tversky approach came under intense attack in the years following the publication of their work and gave rise to the rationality debate (Cohen, 1981; Einhorn and Hogarth, 1981; Lopes, 1991; Gigerenzer, 1991, 1993, 1994; Gigerenzer and Hoffrage, 1995).

This debate is often misplaced and highly muddled for, at least, the following three reasons.[6]

First, there is a basic misunderstanding of the positions of the protagonists. Kahneman–Tversky show that individuals fall short of the unambiguous statistical benchmarks required in the rational actor model in economics. Gigerenzer–colleagues show that in some domains, judgment heuristics perform quite satisfactorily against selected statistical benchmarks. Of course, both Kahneman–Tversky and Gigerenzer–colleagues may be right, and probably are right.

Second, it has been argued that data presented in terms of natural frequencies as compared to probabilities reduces or eliminates biases relative to the best statistical benchmark (Gigerenzer et al., 1988; Gigerenzer, 1991, 1993, 1994; Gigerenzer and Hoffrage, 1995). But this position is vastly overstated (Kahneman and Tversky, 1996). Even in the best case scenarios, 60% or more of the subjects continue to exhibit base rate neglect, across studies, in the presence of frequency information (Barbey and Sloman, 2007). Even for the conjunction fallacy that has been held up as the leading example of the facilitating role of the frequency format, a between-subjects design reveals a high degree of conjunction bias (Kahneman and Tversky, 1996). There are many heuristics such as the anchoring heuristic that are hardly influenced by the choice between frequencies and probabilities.

Third, the high degree of heterogeneity among subjects who exhibit biases contradicts conformity with a single rational response (Stanovich and West, 2000). Finally, there are a large number of other misconceptions about the Kahneman–Tversky position relating to context, framing effects, errors, subject misperceptions, and evolutionary adaptation that have already been well addressed and do not have a firm basis (Kahneman and Tversky, 1996; Stanovich and West, 2000; Kahneman, 2011). Kahneman and Tversky's (1996, p. 584) frustration is well captured in this quote: "The position described by Gigerenzer is indeed easy to refute but it bears little resemblance to ours. It is useful to remember that the refutation of a caricature can be no more than a caricature of a refutation."

[6] We omit here the debate over which mental model (e.g., the dual systems model) best provides foundations for heuristics and biases (Stanovich and West, 2000). These issues are currently more central to psychologists than to economists.

Section 19.16 considers the rationale and implications of attributes of goods/services that might be *shrouded*. Attributes are said to be *shrouded* if they are partly hidden, but can be discovered by individuals at some cost. For instance, while opening a bank account, the minimum balance fees, fees for bounced checks, or other bank surcharges, might not have been very visible in the information provided. There are no advantages to firms from shrouding information in the classical case with fully rational consumers and competitive firms (Jovanovic, 1982; Milgrom, 1981). However, one does observe many examples of shrouded attributes. For instance, charging individuals high shipping costs (shrouded attribute) but low base prices (unshrouded attribute) on eBay increases sales and revenues (Hossain and Morgan, 2006). We follow up the empirical evidence with a formal model of shrouding of attributes (Gabaix and Laibson, 2006). In the presence of myopic consumers who pay insufficient attention to shrouded attributes, it may be optimal for firms to shroud information.

Section 19.17 considers the implications of *limited attention* that reflects limited cognitive abilities. Classical economics assumes full attention, which is robustly rejected by the evidence. The problem of limited attention is relatively more serious for the poor whose attention is diverted by basic problems of food, shelter, and clean water (Datta and Mullainathan, 2014; WDR, 2015). This diversion of attention is a form of *cognitive tax* (Shah et al., 2012; Mullainathan and Shafir, 2013). It reduces the quality of attention towards other important decisions; this is sometimes known as the problem of *limited bandwidth* (Banerjee and Mullainathan, 2008). The WDR (2015) is an excellent introduction to these issues.

In other contexts, consumers pay less attention to taxes that are not listed on price stickers but added at cash registers; displaying information about taxes on price stickers reduces sales (Chetty et al., 2009). Individuals pay limited attention to the rules of earned income tax credit (Hotz and Scholz, 2003). People pay less attention to electronic toll collections, as compared to manual toll collections, allowing operators to levy higher tolls (Finkelstein, 2009). Car sales in used markets are subjected to a left digit bias in odometer readings. Thus, car sales and car prices are sensitive to integer multiples of the 10,000 mile threshold (Lacetera et al., 2012; Englmaier et al., 2014). We discuss the recent work on the *sparse max operator* that formalizes a way of providing differential attention to various attributes of a good (Gabaix, 2014). We close with brief comments on the rational inattention models (Sims, 1998, 2003).

Section 19.18 collects a range of empirical results to show that experts exhibit biases in the same manner as non-experts (Tetlock, 2002; Ben-David et al., 2013; WDR, 2015). Hence, market experience does not eliminate non-standard behaviors.

## 19.2 The law of small numbers

In statistics, the *weak law of large numbers* is often expressed as follows.

> **Theorem 19.1** *(Weak Law of Large numbers; Hogg et al., 2005) Let* $\{X_n\}$ *be a sequence of independently and identically distributed random variables, with a common mean* $\mu$ *and finite variance* $\sigma^2$*. Denote the sample mean based on n observations by* $\overline{X}(n) = \sum_{i=1}^{n} X_i/n$*. Then* $\overline{X}(n)$ *"converges in probability" to the population mean,* $\mu$*, i.e.,* $\lim_{n\to\infty} P\left[\left|\overline{X}(n) - \mu\right| \geq \varepsilon\right] = 0 \; \forall \; \varepsilon > 0$.

The weak law of large numbers can be strengthened to the *strong law of large numbers,* which guarantees almost sure convergence of the sample mean to the population mean (Chung, 1968). Notice that the law of large numbers is a limiting result and holds only for very large sample sizes.

By contrast, individuals are said to subscribe to the *law of small numbers* if they believe that *for any sample size*, the sample mean is identical to the population mean. Such people believe that the sample proportions mimic the population proportions for any sample size.

Consider the following experiment reported in Tversky and Kahneman (1974). A town is served by two hospitals. In the larger hospital, about 45 babies are born each day, and in the smaller hospital, about 15 babies are born each day. As you know, about 50% of all babies are boys. However, the exact percentage varies from day to day. For a period of 1 year, each hospital recorded the days on which more than 60% of the babies born were boys. Which hospital do you think recorded more such days?

In the sample 53 students said that both hospitals are equally likely to have recorded such days and an equal split of 21 students chose the larger and the smaller hospital, respectively. The correct answer is the smaller hospital. The reason is that the larger hospital has a larger sample size, so it is likely to better reflect the population proportion of 50%.

### 19.2.1 *Representativeness heuristic*

Tversky and Kahneman (1974) explain the *representativeness heuristic* as follows. "What is the probability that object A belongs to class B? What is the probability that event A originates from process B?. . . In answering such questions, people typically rely on the representativeness heuristic, in which probabilities are evaluated by the degree to which A is representative of B, that is, by the degree to which A resembles B."

In one experiment, they give a description of a hypothetical individual named Steve (p. 1124): "Steve is very shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail." Subjects in experiments were then asked to judge the probability of various professions that Steve might be engaged in, such as farmer, physician, librarian, salesman, and airline pilot. Subjects in experiments typically judged Steve's profession by matching his description with their stereotype of that profession (representativeness). However, people do not take account of the fact that there are many more farmers than librarians; this is base rate neglect that is considered in greater detail, below.

Consider a more formal example of the representativeness heuristic (Camerer, 1987, 1990, 1995). Suppose that there are two urns, $i = A, B$, and one is chosen randomly by nature. Urn A has 1 red and 2 black balls. Urn B has 2 red and 1 black ball. It is common knowledge that nature chooses urn A with probability 0.6, so $P(A) = 0.6$ and $P(B) = 0.4$. A sequence of three balls is drawn *with replacement* from one of the urns. Subjects in experiments do not know which urn the balls are drawn from.

Let $x$ denote the number of red balls that come up in the sample of three balls, $x = 0, 1, 2, 3$. Suppose that the sample, based on three draws, turns out to be $x = 1$. What is the posterior probability that the sample came from urn A? Using Bayes' rule, this can be computed as follows.

$$P(A \mid x = 1) = \frac{P(x = 1 \mid A)\, P(A)}{P(x = 1 \mid A)\, P(A) + P(x = 1 \mid B)\, P(B)}. \tag{19.1}$$

In a single draw of a ball from an urn, there are two possibilities (red ball or black ball), hence, this is a Bernoulli experiment. Let $p_A, p_B$ be the respective probabilities of drawing a red ball from urns A and B. Then, the thrice repeated Bernoulli experiment has a binomial distribution given by

$$P(x \mid i) = \left( \begin{array}{c} 3 \\ x \end{array} \right) p_i^x (1 - p_i)^{3-x}; i = A, B \text{ and } x = 0, 1, 2, 3. \tag{19.2}$$

A simple calculation shows that $P(x = 1 \mid A) = 4/9$ and $P(x = 1 \mid B) = 2/9$. Thus, from (19.1), we get the Bayesian estimate, $P(A \mid x = 1) = 0.75$. In contrast to the Bayesian estimate, an individual who uses the representativeness heuristic draws one of the following inferences.

1a. $P(A \mid x = 1) = 1$. In this interpretation, the (small) sample has exactly the same proportion of balls of each color as urn A. Hence, representativeness leads individuals to infer with certainty that the sample came from urn A.

1b. $P(B \mid x = 2) = 1$. As in 1, if the sample comprises of 2 red balls and 1 black ball, then representativeness guides the individual to infer with certainty that the balls came from urn B.

2. $0.75 < P(A \mid x = 1) < 1$. Here, the individual makes an inference that is intermediate between a Bayesian posterior and the extreme form of the representativeness heuristic in 1a.

## 19.2.2 *Gambler's fallacy and hot-hands fallacy*

Evidence for the law of small numbers comes from at least three kinds of tasks. In *production tasks*, individuals are asked to produce random sequences. In *recognition tasks*, subjects are asked to recognize if the given sequence is random. In *prediction tasks*, subjects are asked to predict the next element in a sequence.

### INABILITY TO PRODUCE A SEQUENCE OF RANDOM NUMBERS

One implication of the law of small numbers is that many people are unable to generate a truly random sequence of events. Suppose that individual subjects are asked to generate a sequence of random tosses of a coin; denote heads by $H$, and tails by $T$. Most experimental studies find that subjects switch the outcomes too frequently, *as if* to preserve the population proportions (50–50), even in a small sample. So, for instance, the sequence $H, H, H, H, H$ is very likely to be followed by $T$. In other words, the supposedly random sequences generated by individuals exhibit too much negative autocorrelation (Bar-Hillel and Wagenaar, 1991).[7]

### GAMBLER'S FALLACY

If subjects produce *negative autocorrelation* when asked to produce a hypothetical random process, they are said to engage in the *gambler's fallacy*. In a well-known example, there is reduced betting on a winning number in subsequent lottery draws (Clotfelter and Cook, 1993); subjects using the law of small numbers mistakenly think that the winning number has a lower probability of coming up in subsequent draws, despite an unchanged objective probability.

Table 19.1 shows the results of a production task in Rapoport and Budescu (1997) reported by Rabin and Vayanos (2010). These subsequences are culled from an underlying hypothetical sequence of 150 tosses produced by subjects in experiments. The probabilities in the last column are the fraction of subjects who make a choice of H, conditional on the subsequence in the three leftmost columns. On average, the probability of H on the next flip, given that the last toss resulted in a H is $\frac{30+38+41.2+51.3}{4} = 40.1$. It follows, that conditional on the last toss resulting in a H, the average probability of T on the next toss is $100 - 40.1 = 59.9$; this is sometimes

[7] For important empirical work that served to further establish this phenomenon, see Rapoport and Budescu (1992, 1997) and Budescu and Rapoport (1994).

**Table 19.1** Assessed probability that the next flip of a coin will be heads (H), given a sequence of three flips; each flip is either heads (H) or tails (T). Based on Rapoport and Budescu (1997, Table 7, p. 613).

| 3rd-to-last | 2nd-to-last | Very last | Prob. next will be H (%) |
|---|---|---|---|
| H | H | H | 30.0 |
| T | H | H | 38.0 |
| H | T | H | 41.2 |
| H | H | T | 48.7 |
| H | T | T | 62.0 |
| T | H | T | 58.8 |
| T | T | H | 51.3 |
| T | T | T | 70.0 |

Source: Rabin and Vayanos (2010).

referred to as the *switching rate*. The difference between the probabilities of T and H is 19.8%, when the objective difference should be 0%. This negative autocorrelation suggests a gambler's fallacy.

There is negative autocorrelation not only with respect to the very last coin flip but also with respect to the "2nd-to-last" flip also. Changing the outcome from H to T on the "2nd-to-last" flip raises the average probability of an H on the next toss (last column in Table 19.1) from 43.9% to 56.1%, an increase of 12.2%. Similarly, changing the outcome from H to T in the "3rd-to-last" flip raises the average probability of an H in the next toss from 45.5.% to 54.5%, an increase of 9%. Rabin and Vayanos (2010) propose a model that can account for the gambler's fallacy when the outcomes are not binary and are not iid.

A great deal of the evidence for the gambler's fallacy has come from the study of betting behavior. Clotfelter and Cook (1993) and Terell (1994) find that betting on the winning number falls in pick-3 and pick-4 games and takes several months to recover. Metzger (1984) found that people bet lower amounts on long-shot horses if another long-shot horse had won earlier in the day (the two horses can be different horses). Terrell and Farmer (1996) and Terrell (1998) show that betters underbid on a previous race winning number in dog races, even if different dogs are involved in the two races.

Croson and Sundali (2005) confirmed the gambler's fallacy from real-world data obtained from casino players with real stakes. They considered betting on roulette wheels where one can either engage in *inside bets* on numbers 0–36 or on *outside bets* (e.g., on even/odd numbers, on numbers shaded red or black, on numbers 1–18, or 1–12, and so on). The method of identifying the gambler's fallacy is to observe a streak of winning numbers, then check if there is reduced betting on these numbers in subsequent bets. For instance, betting on Black following a streak of outcomes of color Red is an example of the gambler's fallacy. Streaks of length $l = 1, 2, \ldots, 6$ are considered.

Figure 19.1 shows the results for streaks of lengths 1–6 for all outside bets. For instance, 531 bets were placed on one of the outside bets after observing one spin of the roulette wheel that produced an outside outcome; 255 of these bets, or 48%, of these bets satisfied the criterion of the gambler's fallacy. Following streaks of length 2, 258 outside bets were placed, of which the gambler's fallacy was observed in 51% of cases. Thus, for streaks of small length, the observed

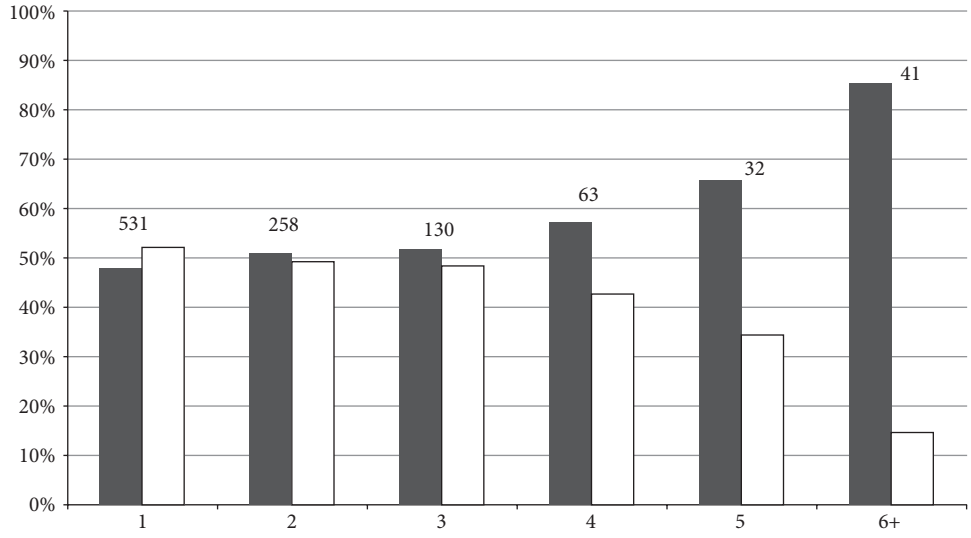


**Figure 19.1** Results for streaks of length 1–6 for all outside bets. The height of shaded bars is the percentage of outside outcomes and the height of blank bars is the percentage of outside bets. The horizontal axis measures streaks of length $l = 1, 2, \ldots, 6$.

Source: With kind permission from Springer Science + Business Media: *Journal of Risk and Uncertainty* 30(3): 195–209. Croson, R. and Sundali, J. "The gambler's fallacy and the hot hand: empirical data from casinos." © 2005 Springer Science + Business Media.

percentages are no different from random. However, statistically significant effects of gambler's fallacy are found for streaks of length 5 and 6. For instance, for $l = 6$, 85% of the bets are consistent with the gambler's fallacy.

The gambler's fallacy has also been applied to finance. For instance, consider the *disposition effect*, which is the tendency to sell winning stocks and hold on to loss making stocks for too long (Shefrin and Statman, 1985; Odean, 1998). The idea is that people think that a stock that has risen in the past is now due for a fall and vice versa, which is consistent with the gambler's fallacy. Odean (1998) found that people sold winning stock at a 50% higher rate than losing stock. This is a robust phenomenon and documented widely for Finnish investors (Grinblatt and Keloharju, 2001), for Israeli investors (Shapira and Venezia, 2001), for Chinese investors (Feng and Seasholes, 2005), for employee stock options (Heath et al., 1999), and for trading volumes following IPOs (Kaustia, 2004; Brown et al., 2006). The robustness of the disposition effect can be gauged from the fact that even when attempts are made to debias the subjects by making the purchase price very salient in experiments at the time of the selling decision, the disposition effect is reduced by only 25% (Frydman and Rangel, 2014).

Furthermore, while the disposition effect is relatively stronger for individual investors, institutional investors are also subjected to the disposition effect (Frazzini, 2006; Chen et al., 2007; Choe and Eom, 2009). However, Barber et al. (2007) paint a slightly nuanced picture for Taiwanese data; corporate investors and dealers exhibit the disposition effect but mutual funds and foreign investors do not. There is some evidence that the disposition effect can be mitigated with experience (Feng and Seasholes, 2005; Seru et al., 2010), and with the level of wealth (Dhar and Zhu, 2006); and it is more prevalent for hard to value stocks (Kumar, 2009).

While prospect theory has been used to justify the disposition effect, the predictions of the relevant theoretical models are sensitive to the auxiliary assumptions used.[8] It is more straightforward to explain the disposition effect using the gambler's fallacy. Stocks that are doing well in the past are believed to revert to lower returns, so individuals sell them too early. Stocks that are not doing well are believed to appreciate, so investors hold on to them for too long.

### HOT HANDS FALLACY

In contrast to the gambler's fallacy, there are instances where subjects may come to believe that there exists *positive autocorrelation* in an inherently random process. The inference of such positively autocorrelated streaks, known as the *hot hands fallacy*, has been documented in many contexts, such as in the game of basketball.

When basketball players are identified to be on a hot streak, observers assign a high probability that they will be successful in making the next shot too. However, the data does not always justify the assignment of a hot streak, given the small sample at hand. Indeed, several pioneering studies showed that the observed sequence of successful basketball shots was statistically a random sequence (Gilovich et al., 1985; Tversky and Gilovich, 1989a,b). In contrast, the betting behavior of individuals reflects beliefs in such streaks (Camerer, 1989).

Recent, unpublished research, has raised the possibility of actual hot streaks in basketball and baseball, thus, there might be no hot hands fallacy in these contexts (Green and Zwiebel, 2013; Bocskocsky et al., 2014; Miller and Sanjurjo, 2015). The basic idea is that in games like basketball, a player on a hot streak is likely to be more tightly marked by the opposition, and forced to take more difficult shots, hence, reducing the observed impact of hot streaks. Controlling for these factors, a weak hot hands phenomenon has been found in some studies. In Bocskocsky et al. (2014), for instance, players who have hot hands are 1.2 to 2.4 percentage points more likely to make a successful shot in basketball. As the authors note, there are several potential issues with how the level of difficulty of a shot is described, as well as the construction of non-standard measures (such as "complex heat") used to infer the extent of hot hands.

By contrast, Miller and Sanjurjo (2015) found stronger evidence of hot hand effects in the neighborhood of 7 percentage points. It should be pointed out, lest it not be obvious, that a statistical demonstration of the hot hands phenomenon does not necessarily negate the hot hands fallacy. For the negation to occur, it would be necessary to demonstrate that individuals see through the statistical occurrence of the hot hands phenomenon. For instance, a control group of subjects could be tested on a randomly generated sequence of basketball shots with streaks in them. Can the subjects spot that the underlying process is purely random? To the best of my knowledge such controls have not been carried out yet, hence, claims that the hot hands fallacy is not real are premature.

However, there is good evidence from other contexts where the *hot hands phenomenon* does not exist, yet the *hot hands fallacy* exists. In a novel field experiment, Guryan and Kearney (2008) find that sales at lotto stores that have sold a winning ticket soar in the immediate weeks following the lotto win; Figure 19.2 plots the log sales at lotto stores in the 40 previous and 40 subsequent weeks following a winning ticket sold at the store. The effect does diminish over time, yet it lasts for an impressive 40 weeks following the win. One possible confounding factor is that a lottery win in a zip code makes the act of playing a lottery salient. In order to check for this, the sales figures of all lotto stores in the winning zip code area are also examined. There is some overall

[8] For prospect theory explanations of the disposition effect, see Barberis and Xiong (2009), Hens and Vleck (2011), Henderson (2012), and Kaustia (2010).

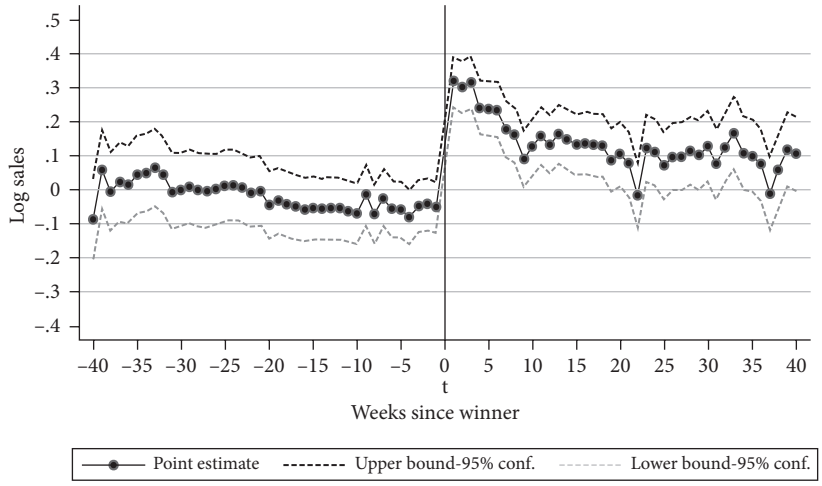


**Figure 19.2** Retail level data for Lotto Texas wins.
Source: Guryan and Kearney (2008) with permission from the American Economic Association.

increase in sales in the same zip code, suggesting an advertisement, or spillover, effect. However, relative to all stores within the zip code, or stores within a mile of the winning store, the sales increase in the winning store is significantly higher, suggesting a store-specific effect.

A natural explanation for the increased sales at winning stores is the hot hands fallacy. In this case, the hot hands fallacy does not arise from a winning streak, but rather from a single observation. However, bettors typically observe a streak of losses and a win at one store, while there is a continuing streak of losses at all other stores. This observation suggests a possible expansion in the scope of the hot hands fallacy. A competing explanation is that people believe in non-existent variation in luck across stores.

The effect on sales in the winning store is increasing in the size of the win. A week after the win, sales increased in the Texas Lotto stores by 37.7%. The corresponding increase in sales for smaller jackpot games is 21.5% (for the Texas Two Step) and 12.4% (for the Cash Five game). Furthermore, the effect is stronger in areas that have more high school dropouts, higher rates of poverty, and a greater proportion of elderly people, suggesting that some or all of these factors may enhance cognitive biases that give rise to the observed effect.

The hot hands fallacy is also found in Croson and Sundali (2005) who study gambling on roulette wheels. Here, 80% of the subjects quit playing after losing on a spin. In contrast, only 20% quit on winning. This is consistent with the idea that winning players wish to continue playing because they believe that luck is on their side (i.e., they are *hot*, or on a *roll*). The regression results confirm this finding. In this case, for individual-level data, there is a statistically significant effect of winning in one round and the number of bets placed in the next round after controlling for individual heterogeneity.

Jørgensen et al. (2011) find the presence of the gambler's fallacy and the hot hands fallacy using panel data from the Danish State Lottery. There are two types of individuals in their data. The first type picks the same numbers every week. The second type of individuals avoids the non-winning numbers they chose last time (gambler's fallacy) and are more likely to choose numbers that are perceived to be on a winning streak (hot hands fallacy).

Powdthavee and Riyanto (2015) show that individuals are willing to pay for an expert who they believe has a hot hand in predicting inherently random sequences of numbers. This theme is

explored in chapter 20 in Kahneman (2011) under the title "The Illusion of Validity." He argues that the returns to professional funds are typically (although not always) the result of luck rather than skill.[9] As evidence, he cites the fact that the temporal correlations in performance for any mutual fund are quite weak, and close to zero, which is consistent with the luck story. Yet, small investors are willing to pay significant amounts for expert financial advice. Malmendier and Tate (2009) find that superstar CEOs (based on previous performance) underperform in subsequent periods. This is also consistent with the story in which high outcomes are achieved, at least partially, by luck, but one eventually experiences regression to the mean (see Section 19.10.1). Some of these issues are taken up in greater detail in Section 21.7 in Chapter 21.

### WHY DO THESE FALLACIES ARISE?

Tversky and Kahneman (1974) report that in successive tosses of an unbiased coin, the sequence $H, T, H, T, T, H$ is considered more likely than the sequence $H, H, H, T, T, T$; however, both sequences have identical statistical probabilities. One plausible explanation is that the second sequence is not considered to be *representative* of the population proportions. As the authors beautifully put it (p. 1125): "Chance is commonly viewed as a self-correcting process in which a deviation in one direction induces a deviation in the opposite direction to restore the equilibrium. In fact, deviations are not 'corrected' as a chance process unfolds, they are merely diluted." This also explains the role of the representativeness heuristic in the gambler's fallacy; gamblers do not bet on the previously winning numbers because they believe that deviations will be corrected, rather than diluted over a sufficiently long period of time.

Like the gambler's fallacy, the seemingly opposite phenomenon of positive correlation when the underlying process is the hot hands fallacy, can also be potentially explained by the representativeness heuristic. In this case, following several identical signals, the sequence of observed signals appears more and more representative of a positively autocorrelated underlying random process. An observer who is subject to the hot hands fallacy then makes predictions based on the incorrectly perceived underlying random process.

Ayton and Fischer (2004) suggest that the domains of the gambler's fallacy and the hot hands fallacy might be different. They propose that the hot hands fallacy is more likely to arise for human performance, as is the case with basketball and tennis players, rather than for inanimate processes such as throws of a dice or coin tosses. The reason is that a streak in sports may signal unobservable variables about a player to observers, such as the level of confidence, fatigue, and the quality of training prior to the game. These unobservable variables do not apply to inanimate processes for which subjects may switch too often in small samples to preserve the population proportions, as suggested by the representativeness heuristic. This also potentially explains emerging evidence of hot hands in basketball and baseball referenced above, yet in inanimate processes such as the outcome of lotteries, there is strong evidence of the hot hands fallacy (see above).

In one set of experiments, the authors presented the subjects with various sequences of outcomes that had varying degrees of positive and negative autocorrelation. Subjects were asked to guess the source of these sequences: human skilled performance (e.g., a professional basketball player's scoring attempts, or a professional tennis player's first serve in) or an inanimate process (e.g., a coin toss, or a binary outcome roulette wheel). Each sequence consisted of 21 binary

[9] Kahneman writes (2011, p. 218): "The idea that large historical events are determined by luck is profoundly shocking, although it is demonstrably true." He gives an interesting example. Before an embryo is fertilized there is a 50–50 chance that it will be a male or a female. Thus there was a 1/8 chance that the world could have survived the three dictators, responsible for a large number of twentieth century deaths—Hitler, Stalin and Mao Zedong."

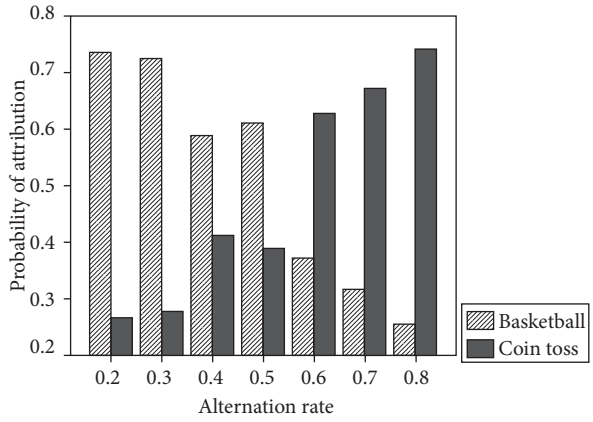


**Figure 19.3** Is the observed sequence with a given alternation rate more likely to have come from basketball or a coin toss?

Source: With kind permission from Springer Science + Business Media: *Memory and Cognition* 32: 1369–78. Ayton, P. and Fisher, I. (2004). "The hot hand fallacy and the gambler's fallacy: two faces of subjective randomness?" © 2004 Springer Science + Business Media.

outcomes. The alternation rate is the number of times a sequence changes direction divided by 10.

The results for one of the comparisons is shown in Figure 19.3. Consistent with the authors' predictions, sequences with low rates of alternation were rated by subjects to have more likely come from basketball rather than coin tosses. Conversely, sequences with high rates of alternation were rated as more likely to have come from a coin toss. These results highlight when the gambler's fallacy is more likely to arise (high alternation rates arising from an inanimate process) and when the hot hands fallacy is more likely to arise (low alternation rates and involvement of human skills).

### 19.2.3 *A formal model of the law of small numbers*

Consider one possible method of formalizing the law of small numbers in a simple model with the following features (Rabin, 2002).[10]

1. Move by nature: Nature chooses a parameter $\theta \in \Theta$. Individuals have a prior distribution $\pi(\theta)$ over the set $\Theta$.
2. The *actual* data generating process: Consider a Bernoulli trial with two possible outcomes: $x$ (success) and $y$ (failure) that occur with respective probabilities $\theta \in [0,1]$ and $1-\theta$. Suppose that we independently repeat the Bernoulli trial $n$ times, with replacement. Denote by $n_x$, the number of times that the outcome equals $x$ in the $n$ trials. It is well known that the number of successes follows a binomial distribution; the probability that there are $n_x$ successes out of $n$ Bernoulli trials is given by

$$p(n_x) = \binom{n}{n_x} \theta^{n_x}(1-\theta)^{n-n_x}; \;\; n_x = 0, 1, \ldots, n.$$

[10] For a theoretical application to finance, see Section 21.6 below.

3. Misperception about the data generating process: The true underlying process is one with replacement. An individual who knows the true model of the world is known as a *Bayesian*. However, suppose that individuals believe, incorrectly, that the data is generated by the following process, without replacement.
   Let time period $t$ be an odd number. At time $t$, an urn has $N$ balls; $\theta N$ balls denote outcome $x$ and the remaining, $(1-\theta)N$, balls denote outcome $y$. A ball is drawn with replacement at time $t$. At time $t+1$, an even time period, the individual believes (incorrectly) that there are only $N-1$ balls left in the urn. In period $t+2$, which is an odd time period, the urn is replenished again to $N$ balls.
   Thus, in an odd period, the probability of outcome $x$ is $\frac{\theta N}{N}$. Now suppose that outcome $x$ materializes in an odd period. Then, in the following period (an even period), the individual believes (incorrectly) that the probability of the outcome $x$ is $\frac{\theta N-1}{N-1}$, while the probability of outcome $y$ is $\frac{(1-\theta)N}{N-1}$. To ensure that the problem is not vacuous, we assume that there are at least two balls corresponding to each outcome in every possible urn. So $\theta N \geq 2$ and $(1-\theta)N \geq 2$.
   Such an individual, who believes in an incorrect model of the world, but is otherwise skilled in all aspects of classical statistical inference, is referred to as an *N-Freddy*. If $N$ is very large, then $N$-Freddy behaves very much like a Bayesian but for small $N$ he can be very biased.

### CERTAINTY ABOUT THE TRUE $\theta$ AND THE GAMBLER'S FALLACY

The gambler's fallacy is a direct implication of the model. Consider the probability of outcome $x$ in an "even period" following outcome $x$ in the previous (odd) period, i.e., the conditional probability $P(x \mid x)$. For a Bayesian who uses the correct model, $P(x) = P(x \mid x) = \theta$. In contrast, for $N$-Freddy, who uses an incorrect model, $P(x) = \frac{\theta N}{N} = \theta$ and $P(x \mid x) = \frac{\theta N-1}{N-1} < \theta$. Hence, following an outcome $x$, $N$-Freddy reduces the probability that the next draw will produce the signal $x$, which is precisely the gambler's fallacy. In contrast, the Bayesian is not subject to the gambler's fallacy.

Suppose we say that Freddy gets even "Freddier" if $N$ falls. What are the implications for the gambler's fallacy? For an $N$-Freddy, $\frac{\partial P(x|x)}{\partial N} = \frac{1-\theta}{(N-1)^2} > 0$. So, as $N$ falls, $P(x \mid x)$ also falls and the gambler's fallacy becomes even worse.

Whilst the Bayesian decision maker and Freddy might revise beliefs in the same direction, there are systematic differences between their predictions. To see this, suppose that there are two values of $\theta : \theta_L < \theta_H$. Starting in an odd period, consider a sample of two consecutive $x$ outcomes. For any $\theta$, the likelihood of this sample for Freddy is $L(x, x \mid \theta) = \frac{\theta N}{N} \frac{\theta N-1}{N-1} = \frac{\theta(\theta N-1)}{N-1}$. Thus for two different values, $\theta_L, \theta_H$, the likelihood ratio for Freddy is

$$\phi_F = \frac{L(x, x \mid \theta_H)}{L(x, x \mid \theta_L)} = \frac{\theta_H(\theta_H N - 1)}{\theta_L(\theta_L N - 1)}. \tag{19.3}$$

On the other hand, for a Bayesian, who uses the correct model of the world, $P(x) = P(x \mid x) = \theta$, so the relevant likelihood ratio is

$$\phi_B = \frac{L(x, x \mid \theta_H)}{L(x, x \mid \theta_L)} = \frac{\theta_H^2}{\theta_L^2}. \tag{19.4}$$

Comparing (19.3) with (19.4) we get that $\phi_F > \phi_B$. Thus, Freddy assigns a lower likelihood to the signal coming from $\theta = \theta_L$ relative to the Bayesian. In other words, observing two successive $x$ outcomes, Freddy assigns a much greater likelihood of the urn containing a larger number of $x$ outcomes, relative to a classical statistician.

### UNCERTAINTY ABOUT THE TRUE $\theta$ AND OVERCONFIDENCE

Now suppose that the urn has 4 balls, $N = 4$ (so we have a 4-Freddy) and Freddy is uncertain about the true value of $\theta$. Suppose that the outcome can either be *good*, $x$, or *bad*, $y$. One may imagine that the draw of a ball from an urn corresponds to the outcome of a production/financial task that is overseen by a manager. We make two assumptions.

1. The probability of the good outcome, $\theta$, takes three equally likely values, $\theta_L = \frac{1}{4}, \theta_M = \frac{2}{4}$, $\theta_H = \frac{3}{4}$. Thus, $\theta$ indexes the competence of the manager; higher values denote greater competence.
2. The urn is now replenished every four periods and the starting period is odd as before.

Thus, a 4-Freddy believes that the outcomes for a type $\theta_L$ manager are drawn from an urn containing four balls such that only one ball corresponds to the good outcome $x$ and three balls correspond to the bad outcome $y$, thus, $x = 1, y = 3$. Similarly, for types $\theta_M$ and $\theta_H$, $x = 2$ and $x = 3$, respectively. On the other hand, a Bayesian knows that all draws are independently drawn from a Bernoulli distribution and the urn is replenished every period.

Suppose that starting in period 1, the performance of the manager is good for two consecutive periods. The Bayesian knows that the random draws are identical and independent, so for any $\theta$, he assigns the conditional probability $P(x, x \mid \theta) = \theta^2$, and, in particular,

$$P(x, x \mid \theta_L) = \frac{1}{16}, P(x, x \mid \theta_M) = \frac{1}{4}, P(x, x \mid \theta_H) = \frac{9}{16} > \frac{1}{2}. \tag{19.5}$$

Given assumptions 1 and 2, the 4-Freddy instead assigns the conditional probabilities (or sample likelihoods) as follows.

$$P(x, x \mid \theta_L) = \frac{1}{4} \times \frac{0}{3} = 0, P(x, x \mid \theta_M) = \frac{2}{4} \times \frac{1}{3} = \frac{1}{6}, P(x, x \mid \theta_H) = \frac{3}{4} \times \frac{2}{3} = \frac{1}{2}. \tag{19.6}$$

For any $\theta$, relative to a Bayesian, the 4-Freddy assigns a lower value of $P(x, x \mid \theta)$. The reason is that once an $x$ outcome is obtained, Freddy believes that the urn contains one less $x$ outcome, reducing the probability that a subsequent outcome will be $x$.

We are now interested in computing $P(\theta_H \mid x, x)$, i.e., the posterior belief that the analyst is of the highest type, following a sequence of two good performances. Using Bayes' rule we get that

$$P(\theta_H \mid x, x) = \frac{P(x, x \mid \theta_H)P(\theta_H)}{P(x, x \mid \theta_L)P(\theta_L) + P(x, x \mid \theta_M)P(\theta_M) + P(x, x \mid \theta_H)P(\theta_H)}. \tag{19.7}$$

Noting that $P(\theta_L) = P(\theta_M) = P(\theta_H) = \frac{1}{3}$ (because each type is equally likely) and substituting (19.5) and (19.6) sequentially into (19.7), Freddy computes $P(\theta_H \mid x, x) = \frac{21}{28}$, while a Bayesian computes $P(\theta_H \mid x, x) = \frac{18}{28}$. Thus, based on a sample of two observations, the 4-Freddy is more confident than the Bayesian that the fund manager is of the highest type. This is a form of over-

confidence, based on observing very small samples, which plays an important role in behavioral finance.

Overconfidence is considered in greater detail in Section 21.7 in Chapter 21. Here, we offer some brief comments. Suppose that in financial markets, an observer who behaves like Freddy observes the performances of several financial analysts. Then those analysts who, purely by luck, achieve successive good outcomes, say, $x, x, x$, will be believed by the observer to be more likely to be extremely competent ($\theta = \theta_H$), relative to a Bayesian. Analogously those who achieve a sequence of bad outcomes, say, $y, y, y$, again purely by luck, are believed to be more likely to have low competence ($\theta = \theta_L$), relative to a Bayesian. Indeed, the observer may be willing to pay a premium to use the services of the analyst that they mistakenly perceive to be more competent (Powdthavee and Riyanto, 2015).

This situation is not unrepresentative, even if it is artificially constructed. Kahneman (2011, ch. 20) recounts that he was once invited to speak to a group of financial advisors who provided their services to very wealthy clients. Examining the data on 25 of those advisors over eight years, he ranked the advisors by performance in each year and then computed the correlation coefficients of the ranks in each pair of the eight years. The average of the correlations was close to zero, 0.01. Thus, it appears that it is not skill but luck that governs the relative performance of the advisors and the firm was rewarding luck not skill.

Needless to say, exposure to these results did not go down very well with the analysts. As Kahneman (2011, p. 216) puts it: "Their own experience of exercising careful judgment on complex problems was far more compelling to them than an obscure statistical fact." Since financial analysts are typically good at explaining past financial events with the benefit of hindsight (see hindsight bias in Section 19.8 below), an understanding of the past gives them the mistaken view that they can predict the future.[11] For a theoretical model of overconfidence in finance that is consistent with the views of Kahneman (2011) described above, see Gervais and Odean (2001). In this model, investors who do better than others, purely by luck, view their success as vindication of their skill rather than luck. This creates overconfidence that might nevertheless by tempered by subsequent experience.

## 19.3 Conjunction fallacy

Another illustration of the violation of a basic principle of probability theory is the *conjunction fallacy*. Tversky and Kahneman (1983) posed the following problem to Stanford students who were well-trained in decision sciences.

"Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations." Which of the following statements is most likely?

1. Linda is a bank teller ($B$).
2. Linda is active in the feminist movement ($F$).
3. Linda is a bank teller and is active in the feminist movement ($BF$).

[11] For more details, see Chapters 19 and 20 in Kahneman (2011) under the headings "Illusion of understanding" and "Illusion of validity."

Since the event $BF \subseteq B$ and $BF \subseteq F$, it follows under classical statistics that $P(BF) \leq P(B)$ and $P(BF) \leq P(F)$. Any violation of these inequalities is a *conjunction fallacy*. The problem stated above has three choices: $B$, $F$, and $BF$. An earlier version of the problem was handed out with eight choices. Between 85–90% of the respondents rated the event $BF$ as more likely relative to the separate events $B$ and $F$. Thus, subjects exhibit the conjunction fallacy. In one version, 85% of the sample of doctoral students in the decision-science program at the Stanford graduate school of business committed the conjunction fallacy.

There is a very large literature on the conjunction fallacy. In addition to the literature reviewed below, the interested reader may consult the survey in Moro (2009) and Kahneman (2011, ch. 15). For a skeptical view, see Hertwig and Gigerenzer (1999). For a review of the literature on the causes of the conjunction fallacy, see Tentori et al. (2013).

The *Linda problem*, as it came to be known, was bitterly criticized and continues to be criticized, although, as we show below, no conclusive evidence of its rejection has been given. Indeed Tversky and Kahneman (1983) themselves addressed many potential criticisms, but their caveats were conveniently forgotten.

### 19.3.1 *Does a natural frequency format resolve the conjunction fallacy?*

Hertwig and Gigerenzer (1999) report that if the Linda problem is posed in terms of natural frequencies (e.g., 15 out of 100 rather than 15%), then subjects do not engage in the conjunction fallacy. However, Kahneman and Tversky (1996) find that presenting the information in natural frequencies does not eliminate the conjunction fallacy. They presented the following Linda problem to three groups of subjects (Groups 1, 2, 3) in a natural frequency format.

Linda is in her early thirties. She is single, outspoken, and very bright. As a student she majored in philosophy and was deeply concerned with issues of discrimination and social justice. Suppose there are 1000 women who fit this description. How many of them are

(a) High school teachers? [Groups 1, 2, 3]
(b) Bank tellers? [Groups 1, 2]
(c) Bank tellers and active feminists? [Groups 1, 3]

In the between-subjects treatment, the median response for choice (c) was statistically larger than the median response for choice (b), confirming the conjunction fallacy. Subsequent evidence in Tentori et al. (2004) also supports the view that presenting the information in the form of natural frequencies does not eliminate the conjunction fallacy.

### 19.3.2 *Do monetary incentives and groupthink eliminate the conjunction fallacy?*

Charness et al. (2010) introduce incentives in the Linda problem and allow for consultation in groups of twos (pairs) and threes (trios); as a baseline, they also consider individual decisions (singles). The incentive for giving the correct answer is an extra $2 over the show-up fee of $2. The results are shown in Table 19.2; T&K, 1983 is Tversky and Kahneman (1983) and CLK, 2008 is the working paper that forms the basis of the reported results in Charness et al. (2010).

Tversky and Kahneman (1983) find that 85.2% of their subjects engage in the conjunction fallacy. The corresponding number in Charness et al. (2010), based on a test of proportions, is 58.1%

**Table 19.2** Violations of the conjunction fallacy.

| Study | Details | Incorrect answers/ total sample | Error rate (percent) |
|---|---|---|---|
| *Individuals* | | | |
| T&K, 1983 | UBC undergrads, no incentives | 121/142 | 85.2 |
| CKL, 2008 | UCSB students, singles, no incentives | 50/86 | 58.1 |
| CKL, 2008 | UCSB students, singles, incentives | 31/94 | 33.0 |
| CKL, 2008 | UCSB students, total singles | 81/180 | 45.0 |
| *Pairs* | | | |
| CKL, 2008 | UCSB students, in pairs, no incentives | 27/56 | 48.2 |
| CKL, 2008 | UCSB students, in pairs, incentives | 5/38 | 13.2 |
| CKL, 2008 | UCSB students, total in pairs | 32/94 | 34.0 |
| *Trios* | | | |
| CKL, 2008 | UCSB students, in trios, no incentives | 10/39 | 25.6 |
| CKL, 2008 | UCSB students, in trios, incentives | 5/48 | 10.4 |
| CKL, 2008 | UCSB students, total in trios | 15/87 | 17.2 |

Source: Charness et al. (2010).

and it is statistically smaller. Furthermore, Table 19.2 shows that the violation of the conjunction rule falls significantly when singles are given incentives (from 58.1% to 33%).

The conjunction rule is violated even less when one moves from pairs to trios, as compared to moving from singles to pairs.[12] Thus, "three heads are better than one." In the presence of incentives, trios violate the conjunction rule in only 10.4% of the cases. Since university students are likely to be better trained in probability theory, it will probably be worth replicating these results with a non-student population. Furthermore, the high rates of violations of the conjunction fallacy for singles leaves open the possibility that in real-world situations with singles, the conjunction fallacy may yet play an important role.

In contrast to these results, Bonini et al. (2004) show that monetary incentives do not crowd out the conjunction fallacy. They give subjects the option of allocating 7 euros among three options: (1) Option $B$. (2) Option $B \wedge F$ (where $\wedge$ is the logical operator "and"). (3) Option $B \wedge (\sim F)$, where $\sim F$ is the option "not $F$." The third option controls for the objection that subjects may actually be interpreting (1) as (3), i.e., the option $B$ as $B \wedge (\sim F)$. It is a dominant action to bet on Option $B$, so if subjects did not engage in the conjunction fallacy, we should observe all subjects to bet 7 euros on Option $B$.

Across all blocks of experiments, the average amount allocated to the first option was 1.99 euros. In contrast, the corresponding allocation to options 2 and 3, respectively, the dominated bets, was 3.15 euros and 1.86 euros. The difference between the allocations to options 1 and 2 is statistically significant at 5% significance, using a one tailed t-test. In the five blocks of experiments, the average percentage of subjects that engaged in the conjunction fallacy was 85.32, which is nearly identical to the figure of 85% reported in Tversky and Kahneman (1983). Furthermore,

[12] Similar advantages of group consultation in the reduction of stochastically dominated options is reported in Charness et al. (2007).

the experimental results in Stolarz-Fantino et al. (2003) further argue for the robustness of the conjunction fallacy to monetary incentives and framing effects.

### 19.3.3 *Why does the conjunction fallacy arise?*

One possible explanation of the conjunction fallacy is the use of the representativeness heuristic. The description of Linda as a feminist and a bank teller is relatively more representative of her described traits. Kahneman (2011, p. 158) writes of his initial excitement about the result: "we had pitted logic against representativeness, and representativeness had won!"

Conditional on Linda's description, it is plausible that Linda is a member of the feminist movement. If this were substituted by a non-plausible event (e.g., Linda works for the local council), then there is no presumption that the conjunction fallacy would arise. Consider the following problem given in Kahneman (2011).

Which alternative is more probable?

Mark has hair.

Mark has blond hair.

In this problem, the given information does not make blond hair more plausible, hence, as expected, no conjunction fallacy was found.

Yet another potential explanation of the conjunction fallacy is based on interpreting the role of "and" in the Linda problem (Camerer, 1995, p. 598). The explanation centers on the validity of the interpretation of the word "and" as the logical operator $\wedge$ such that $A \wedge B$ means the intersection of the sets $A$ and $B$. Hertwig et al. (2008) argue that in the statement $A$ and $B$, "and" could be interpreted as (1) events $A$ and $B$ happening in temporal sequence as in "I went to the store and bought some whisky", or (ii) as a causal relationship "Smile and the world smiles with you." Hertwig et al. (2008) then go on to explore the precise sense in which their subjects in experiments interpret "and." They find that subjects who interpret "and" as the logical operator are less likely to be engaged in the conjunction fallacy.

However, Tentori et al. (2004) show that although the conjunction fallacy is lessened by taking account of the literal differences between the conjunction "and" and the logical operator $\wedge$, it nevertheless remains an important fallacy. A fairly detailed discussion of these issues is provided in the careful study by Tentori and Crupi (2012). They concede that the word "and" may have different meanings in different sentences, but within the same sentence its meaning is often unambiguous. Since "and" is one of the most widely used words in the English language, ambiguity of its use within the same sentence would have been disastrous for communication, which has been at the heart of human evolution. They also point out that the multiple conjunctions considered in Hertwig et al. (2008) (e.g., in their experiment 3) have even lower probability than the single conjunctions in the Kahneman–Tversky results and these do not disprove the conjunction fallacy. They argue, in particular, that the theoretical foundations for the claims and the empirical inferences in Hertwig et al. (2008) are questionable. They perform two novel experiments and their results directly contradict the Hertwig et al. results.

Kahneman (2011, pp. 164–5) summarizes his recent feelings about the fallout from his research on the Linda problem: "The Linda problem attracted a great deal of attention, but it also became a magnet for critics of our approach . . . These arguments were sometimes extended to suggest that our entire enterprise was misguided: if one salient cognitive illusion could be weakened or explained away, others could be as well . . . The net effect of the Linda problem was an increase in the visibility of our work to the general public, and a small dent in the credibility of our approach among scholars in the field. This was not at all what we had expected."

## 19.4 The availability heuristic

Tversky and Kahneman (1974) describe the availability heuristic as follows: "There are situations in which people assess the frequency of a class or the probability of an event by the ease with which instances or occurrences can be brought to mind. For example, one may assess the risk of heart attack among middle-aged people by recalling such occurrences among one's acquaintances."

As one might expect, this heuristic can lead to inferential mistakes. In one experiment, Tversky and Kahneman (1974) presented subjects with a mixed-gender list of names in which the men and women were equally represented. In treatment 1, the men in the list were famous but not the women. In treatment 2, the opposite was true. Subjects have to infer if there were more men named in the list than women. In treatment 1 subjects reported a higher number of men, while in treatment 2, they reported more women. Thus, when instances in a class are more easily retrieved, the class appears to be more numerous, even if the alternative classes are equally represented.

Salience of an event might also influence how easily, similar events are retrieved from memory. So, for instance, subjects assign higher probabilities to road accidents or house fires, after having just seen one. Tversky and Kahneman (1973) ask subjects the probability that a randomly picked three letter word from the dictionary begins (event $A$) or ends (event $B$) with the consonants $r$ or $k$. Because it is easier to retrieve words beginning with the letters $r$ or $k$ rather than those that end with these consonants, subjects assigned greater probability to event $A$. However, the correct answer is event $B$.

If subjects cannot recall events from memory, then the ease of hypothetically constructing specific instances will determine which event is thought to be more likely. Tversky and Kahneman (1974) ask subjects in experiments to imagine, or hypothetically construct, all the possible contingencies, risks, or dangers that might arise from a risky expedition. If it is particularly easy to mentally construct these contingencies, then the expedition may be deemed too dangerous and abandoned. Conversely, if the contingencies are too hard to construct, then the opposite could be true. However, ease or difficulty of constructing these contingencies might not be perfectly related to the actual dangers involved in the expedition. Hence, in each case, individuals following such a heuristic could make potentially fatal errors.

More vivid memories may be more easily retrieved. Hence, individuals may assign higher subjective probabilities to a range of events such as airplane accidents, shark attacks, muggings, and snake bites. The media plays a potentially important role in creating vivid memories and selecting for us the subsets of events that can be retrieved. In particular, the media is often driven by considerations of the novelty, rarity, and poignancy of events. In the context of the availability heuristic, Eisenman (1993) shows that, based on media stories, individuals may believe that drug use has increased, when in fact it has decreased. Lichtenstein et al. (1978) find that people's perception of risk is higher for events that are more extensively covered by the media. As Kahneman (2011, p. 138) puts it: "The world in our heads is not a precise replica of reality."

The empirical study of Lichtenstein et al. (1978) has been very influential in establishing the availability heuristic.[13] Their main contribution was to show a tight link between the frequency of death by various causes and two proxies for the availability of such events in human memory. They discovered that the first proxy, *direct experience* (e.g., death of a friend or close relative),

[13] For a review of the empirical literature on the availability heuristic, see Betsch and Pohl (2002) and Kahneman (2011).

predicted better as compared to the second proxy, *indirect experience* (e.g., exposure to statistics via the media). This result is supported by Hertwig et al. (2005) who found that the best predictor of an individual's choice between the risks of two alternative causes of death is the knowledge of similar events in one's social network.

We consider now the empirical study by Pachur et al. (2012) that not only replicates the empirical work of Lichtenstein et al. (1978) but also draws attention to the relation between the availability heuristic and the *affect heuristic* that we consider below in Section 19.5. In the first set of experiments, Swiss university students were given three tasks. In the incentivized *choice task*, among the set of 24 cancers, subjects had to guess for each possible pair (276 pairs in total), the cancer that causes more deaths per year. In the *estimation task*, subjects had to guess the number of deaths caused by each type of cancer per year. In the *valuation task*, individuals were required to make *value of statistical life judgments* (VSL); this required them to place a monetary value, measured in Swiss francs, on one life saved from each type of cancer. For each type of cancer, individuals also reported the number of deaths that they could recall within their social network.

The results are shown in Table 19.3; *Mdn* denotes the median value. For the moment, we ignore column 5, the dread score; it is used later in our discussion of the affect heuristic. The main results are as follows.

1. In the binary *choice task*, there were 62.2% correct answers, i.e., subjects guessed correctly which of the two cancers caused greater fatalities.
2. In the *estimation task*, the correlation between the estimated and actual fatalities is 0.46 ($p = 0.2$).
3. In the *valuation task*, there is much variation in the monetary figures that subjects assign to saving a life from a particular kind of cancer. For instance, the VSL figure on breast cancer is four times higher than on rectal, bladder, or ovarian cancer. A fuller explanation of these results is given in the discussion on the affect heuristic below.
4. The correlation between the median estimated frequency of risk (column 3) and the mean number of recalled instances (column 6) was 0.62 and this is highly significant at the 1% level. The corresponding correlation reported in Lichtenstein et al. (1978) was 0.5 to 0.9.

In a second set of experiments, the authors consider the classic set of risks in Lichtenstein et al. (1978). These risks pertain to diverse causes of death such as diseases, homicides, asthma, suicides, murders, and natural hazards. The three tasks in this experiment (choice, estimation, and valuation) are identical to the first experiment. The main difference is that the last column in Table 19.3 is now split up into direct experiences (number of recalled instances in one's social network) and indirect experiences (the number of recalled instances in the media).

The results are as follows. In the binary choice task that compared the relative mortality rates from two different causes of death, subjects guessed the correct answer in 69% of the cases. The correlation between the estimated frequencies and direct experience is 0.50, which is significant at the 1% level. This suggests an important role of the availability heuristic. However, the correlation of indirect experiences with the estimated frequency is low.

Kuran and Sunstein (1999) introduce the idea of *availability cascades*. This refers to a self-fulfilling sequence of events that are typically sparked by an initial media report about an event that is not major, followed by an emotional, affective, reaction from the public. This gives rise to another heightened round of emotional, self-fulfilling events, and so on. Eventually, and typically, this leads to a policy response that is not justified on cost–benefit terms. The interested reader can consult the authors' explication of their ideas based on two case studies. The *Love Canal affair* and the *Alar scare*.

**Table 19.3** Empirical evidence for the *availability* and *affect* heuristics.

| | Annual mortality rate | *Mdti* frequency estimate | *Mdn* VSL | Affect (M dread score) | Availability (M number of instances in social network) |
|---|---|---|---|---|---|
| Penis cancer | 10.2 | 40.0 | 10000 | 4.65 | 0.00 |
| Testicular cancer | 17.2 | 109.5 | 25000 | 4.55 | 0.03 |
| Bone cancer | 37.5 | 80.0 | 30000 | 4.88 | 0.12 |
| Thyroid cancer | 69.3 | 80.0 | 30000 | 4.44 | 0.03 |
| Larynx cancer | 94.2 | 50.0 | 20000 | 3.92 | 0.24 |
| Cancer of the connective tissue | 94.3 | 45.0 | 17500 | 4.40 | 0.00 |
| Cancer of the gall bladder | 196.5 | 30.0 | 11010 | 4.31 | 0.09 |
| Malignant melanoma (skin cancer) | 242.0 | 50.0 | 17102 | 3.59 | 0.24 |
| Cervical cancer | 295.8 | 95.0 | 20000 | 4.39 | 0.15 |
| Renal cancer | 339.2 | 50.0 | 20000 | 4.30 | 0.03 |
| Cancer of the mouth and throat | 351.0 | 60.0 | 12609 | 3.91 | 0.00 |
| Esophageal cancer | 384.5 | 50.0 | 15000 | 4.31 | 0.09 |
| Rectal cancer | 437.2 | 80.0 | 10000 | 4.14 | 0.00 |
| Bladder cancer | 450.5 | 30.0 | 10616 | 4.23 | 0.03 |
| Ovarian cancer | 453.2 | 100.0 | 20000 | 4.43 | 0.03 |
| Cancer of the nervous system | 455.0 | 90.0 | 30000 | 5.02 | 0.32 |
| Hepatic cancer | 513.0 | 110.0 | 10000 | 4.04 | 0.18 |
| Stomach cancer | 572.2 | 115.0 | 20000 | 4.22 | 0.18 |
| Pancreatic cancer | 897.8 | 60.0 | 17501 | 4.43 | 0.26 |
| Colon cancer | 1172.2 | 120.0 | 10000 | 4.23 | 0.15 |
| Prostate cancer | 1312.3 | 87.0 | 23115 | 4.53 | 0.32 |
| Leukemia and lymphoma | 1331.7 | 100.0 | 35000 | 4.91 | 0.38 |
| Breast cancer | 1347.3 | 200.0 | 40000 | 4.24 | 0.79 |
| Lung cancer | 2756.0 | 300.0 | 22500 | 4.00 | 0.50 |

Source: Pachur et al. 2012.

## 19.5 The affect heuristic

There is a correlation between the ease of retrieval of information about an event, which is the basis of the availability heuristic, and the emotional, or *affective*, reactions to the event. In making judgments, individuals tag the *affect* or the positive and negative quality/feeling *induced* by the representation of an event, signal, or stimulus. These tags, in turn, produce an *affect pool* that individuals draw on, in making judgments and decisions (Slovic et al., 2002). The affective response is the immediate response to a stimulus, as distinct from a deliberative response. We have already encountered the affect heuristic in the book. Two leading examples come from Part 6. These are

the *risk as feelings* hypothesis (Loewenstein et al., 2001) and the work of Antonio Damasio that studied the link between affective reasoning and rational judgment.

In an early contribution, Zajonc (1980) noted that we do not just see a car, we see a *sleek car*, a *cool car*, an *ugly car*, depending on the *affect* of the car on us. Slovic et al. (2002, p. 398) write: "We sometimes delude ourselves that we proceed in a rational manner and weight all the pros and cons of the various alternatives. But this is probably seldom the actual case. Quite often 'I decided in favor of X' is no more than 'I liked X' . . . We buy the cars we 'like,' choose the jobs and houses we find 'attractive,' and then justify these choices by various reasons." This view is radically different from the textbook consumer theory taught to students of economics.

Kahneman (2011) notes that these affective shortcuts are an instance of *substitution* in which one substitutes the answer to a hard question (What do I think about it?) by the answer to an easier question (How do I feel about it?). This provides a cognitive shortcut to engaging in a deliberative response that is time consuming, particularly in the case of complex and difficult problems that must be decided within a time constraint. For this reason, we speak of a heuristic in these situations—the *affect heuristic*.

On the one hand, the affect heuristic offers a speedy solution to complicated problems. On the other hand, it might not lead to an optimal payoff for an individual. Firms routinely use diverse strategies to tap into the affect heuristic of consumers to increase their sales (see the problems at the end of Part 7). The affective system is not very good at tracking small changes, or changes that are remote in time. This explains the difficulty in resisting one more fag (Slovic et al., 2002), which is not optimal for an individual interested in kicking the habit.

The policy recommendation by Cass Sunstein, in several publications, is to base public policy on a strict cost–benefit basis in order to ensure maximum mileage for the taxpayer's dollar. This is in contrast to the position taken by Paul Slovic, the leading advocate of the affect heuristic (see below), who argues that there are no objective risks. Since experts typically measure objective risks (e.g., lives saved or lost) whereas the public has a subjective notion of risk, basing policy on a cost–benefit basis ignores the will of the public, which is not democratic. For a further discussion on this fascinating issue, see Kahneman (2011).

### 19.5.1 *Empirical evidence for the affect heuristic*

Fischhoff et al. (1978) showed that the perception of risks of various hazards depended on the *dread* that these hazards invoked. The hazards included antibiotics, X-rays, vaccines, smoking, and alcohol. Furthermore, the judgment of risks of various hazards, and the subjective benefits that individuals assign to these risks, are negatively correlated. Thus, individuals assign lower benefits to risks that they dread. However, there need not be any relation between the degree of risk and the benefits associated with an event. These results show that human judgments are based not only on what they *think* but also, perhaps mainly, on what they *feel*.

Alhakami and Slovic (1994) found that the relation between perceived risks and benefits from activities was influenced by how *affectively charged* the activity was. Finucane et al. (2000) exploited these results to show that the affect of an event, substance, object, or technology (we use the collective term "hazard") determines both the perceived risk as well as the perceived benefit. Exploiting an implication of this insight they found that (i) the perceived benefits of a hazard can be altered by revealing the degree of risk, and (ii) the perceived risk can be altered by revealing the benefits. For instance, in the case of nuclear power, subjects assigned low (high) risks when it was revealed to them that the benefits are high (low). These results were even stronger when subjects had to make their decisions under greater time pressure.

Slovic et al. (1999) replicated these results with experts who were members of the British Toxicology Association, demonstrating that the affect heuristic is probably hardwired and difficult to undo with training. The experts were given a list of 30 chemical items and asked to make a quick intuitive rating of the items on an *affect scale* that ranged from bad to good; the chemicals included second hand smoking, dioxins in food, benzene, and aspirin. Once the experts had made the affect rating, they were asked to make a risk judgment of exposure to each chemical when the exposure level was 1/100 of the level rated to be dangerous by an appropriate regulatory agency. Clearly, at these levels, the chemicals should carry little or no risk and there should probably be little variation in the assessed risks. However, the affect ranking and the risk ranking of the chemicals was strongly negatively correlated for 95 out of the 97 experts; chemicals that had a relatively bad affect were judged to be relatively more risky.

We now revisit the data in Table 19.3 and interpret the results in column 5 that reports the dread score based on 12 risk characteristics in Slovic et al. (1980). Experts tabulate deaths by objective measures such as the number of lives lost. However, for most individuals, there are *better deaths* and *worse deaths*. For instance, death in a ghastly accident, or a painful death, elicits a more affective response as compared to learning that someone died in their sleep. Thus, the perception of individuals, unlike those of experts, about the risks associated with various cancers, or the risk of death by various causes, are often based on an affective reaction.

For the first set of experiments on the set of 24 different cancers (reported in Table 19.3), the availability heuristic performs better than the dread score in predicting binary choices among fatalities arising from different types of cancers. However, the dread score had better explanatory power in explaining the VSL judgment. The correlation between the two is 0.44 with a p-value of 0.03. In the second set of experiments (classic set of risks in Lichtenstein et al., 1978), the availability heuristic does better than the affect heuristic when it comes to predicting fatalities in binary comparisons of a classic set of risks. However, in the VSL judgment, both heuristics do equally well.

### 19.5.2 *The relative affectiveness of the probability and frequency formats*

Slovic et al. (2000) explored the effect of frequency representations and equivalent probability representations of information on the affect heuristic. The subject pool was experienced forensic psychologists and psychiatrists. Subjects were asked to guess the likelihood that a patient discharged from a mental hospital would commit an act of violence within six months of being discharged. In the first treatment, subjects were given the relative frequency estimate of other experts of similar patients committing the violent act; e.g., 10 out of every 100 such patients commit a violent act. In the second treatment, the view of other experts was given as a percentage (e.g., 10% of similar patients commit a violent act). The frequency estimates of the opinion of other experts (treatment 1) produced significantly higher estimates as compared to the probability estimates (treatment 2). Thus, frequency estimates appear to be more *affective*.

These results are in conformity with the empirical results of Denes-Raj and Epstein (1994). Here, subjects were given a prize for drawing a red bean from two different urns whose composition could be seen in a picture. The majority of the subjects chose the urn with the higher number of red beans (e.g., 10 out of 100) rather than the urn with the higher percentage of red beans (e.g., 2 out of 10). Images of a higher number of red beans are more affective.

## 19.6 Anchoring and adjustment

It is an extremely robust and reliable empirical finding that the inferences made by people are heavily dependent on an *initial suggestion* or *anchor*. The anchor could be informative or entirely uninformative. This phenomenon is known as *anchoring*. Unlike some of the other judgment heuristics, we can numerically measure the precise degree of anchoring. Early studies of the anchoring effect first appeared in psychophysics, which studies the relation between stimulus and sensation. For instance, the perceived duration of a target sound depended on a long or short sound (anchor) that subjects had previously been exposed to; for a brief review of this literature, see LeBoeuf and Shafir (2006).

The field was revolutionized by the work of Tversky and Kahneman (1974), who departed from the practice in psychophysics by focusing on numeric anchors. Their method has since become the established norm for experiments on anchoring. In the typical experiment, subjects are given a pair of questions. The first question asks them to make a *comparative judgment*. Here, subjects indicate if a target quantity, $x_T$, is greater or lower than some anchor, $x_0$. In a second question, subjects are asked for an *absolute judgment*, their best estimate of the target value, $x_T$. Anchoring arises if in the second question, the estimated target value is influenced by the anchor, $x_0$.

### 19.6.1 *Empirical evidence for anchoring*

In a celebrated experiment, Tversky and Kahneman (1974) rigged a wheel of fortune with the numbers 1–100 to stop at either of the two numbers: 10 or 65. Subjects had to write down the number that the wheel stopped at. They were then asked the following two comparative and absolute judgment questions.

*Comparative judgment question*: Is the percentage of African nations among UN members larger or smaller than the number you just wrote?

*Absolute judgment question*: What is your best guess of the percentage of African nations among UN members?

The answers to the second question were found to be anchored too closely on the irrelevant number that came up on the wheel. Those who observed the number 10 (respectively, 65) in the comparative judgment question, answered 25% (respectively, 45%) in the absolute judgment question. Kahneman (2011, p. 120) writes: "We were not the first to observe the effects of anchors, but our experiment was the first demonstration of its absurdity." These results immediately spawned great interest and the results were replicated in a large number of contexts and shed light on many disparate phenomena.

The anchoring effect can be numerically measured. Kahneman (2011) reports the following experiment. Subjects were asked the following two questions in a San Francisco Exploratorium.

1. Is the height of the tallest redwood more or less than $x_0$ feet?
2. What is your best guess about the height of the tallest redwood?

Two different values of the anchor, $x_0$, measured in feet, were given to two different groups of subjects; a high anchor value, $x_0^H = 1200$, and a low anchor value, $x_0^L = 180$. The answers to the target values in the second question were, respectively, $x_T^H = 844$ and $x_T^L = 282$. Using a method proposed by Jacowitz and Kahneman (1995), we can calculate the *anchoring index* as $\frac{x_T^H - x_T^L}{x_0^H - x_0^L} \times 100 = 55\%$.

In Mussweiler and Strack (1999), in the comparative condition, subjects were asked: "Is the mean temperature in Antarctica in winter higher or lower than $-x_0$ °C?" The initial temperature

was intended to serve as an anchor. Two anchors were used, a high anchor, −17 °C, and a low anchor, −243 °C. This is followed by an elicitation of the absolute judgment of subjects of the mean winter temperature in Antarctica. The answer in the absolute judgment condition was closely correlated with the anchor values of $-x_0^\circ$C.

Anchors need not always be provided by the experimenter. They can be self-generated. As an example, consider the experiment in Tversky and Kahneman (1974) in which two groups of subjects was asked to take 5 seconds to find the answers to two calculations: $1 \times 2 \times \ldots \times 8$ and $8 \times 7 \times \ldots \times 1$. Due to the time pressure, the first few numbers are likely to serve as anchors. If this reasoning is correct, then the first calculation gives a low anchor and the second gives a high anchor. The correct answer to the identical calculations is 40,320. However, the median answers to the ascending and descending sequences are, respectively, 512 and 2250, which is consistent with an anchoring explanation.

The finding of Tversky and Kahneman (1974) that implausible anchors influence the outcomes has been replicated widely. Strack and Mussweiler (1997) asked subjects in the comparative judgment question if Mahatma Gandhi was younger than 140 years (high anchor) or older than 9 years (low anchor); both anchors are implausible. However, the respective answers by the two groups of subjects in the absolute judgment question (Mahatma Gandhi's age) were 67 years and 50 years, demonstrating a strong anchoring effect. Mussweiler and Strack (2000a) report similar findings in which subjects had to guess the length of the Mississippi river in the absolute judgment task. In the comparative judgment tasks, two anchors were given. A short but plausible anchor, 2000 miles, and a long, but implausible anchor, 30,000 miles. The answers to the absolute judgment question were significantly influenced by the anchors; mean estimates in the low and the high anchor conditions were 3,768 miles and 12,144 miles.

Field evidence also supports anchoring. Northcraft and Neale (1987) showed that the estimated value of properties by estate agents was influenced by the list price, which served as an anchor. Estate agents were given a 10 page booklet with all relevant information about a property. The booklet also contained a list price, which was either $83,900 (high anchor) or $65,900 (low anchor). Agents exposed to the high anchor priced the property significantly higher and the *anchoring effect* was 41%. Business school students were only slightly more susceptible to anchoring; the anchoring effect in this case was 48%. The estate agents insisted that the anchor had no effect on their calculations, yet the significant anchoring effect is incompatible with their insistence.

Anchoring may be related to entitlements. In Ariely et al. (2006), two groups of students attend a poetry reading session. The first group is asked if they *would pay* $2 to attend the session. The second group is asked if they would attend if they *were paid* $2. The anchor given to the first group suggested that the poetry reading session would give them utility, while the anchor to the second group is suggestive of disutility. Later, both groups were asked if they would attend for free. A third of the subjects in the first group agreed to attend but only 8% in the second group agreed to attend.

In the efficient markets hypothesis (see Chapter 21 for details), the past pattern of prices of securities should not reveal any new information that enables one to predict the future price. Mussweiler and Schneller (2003) use the anchoring heuristic to provide one test of theory. Their subject pool is Business and Economics students at a German university who have had experience of the stock market, ranging from a year to approximately three years. They provide students with background information about a company and the relevant economic conditions, drawn from an actual financial firm.

They also show students either the left or the right panel of Figure 19.4 that plots the stock price of the company over the previous 12 months. The left panel has a salient high price and the right

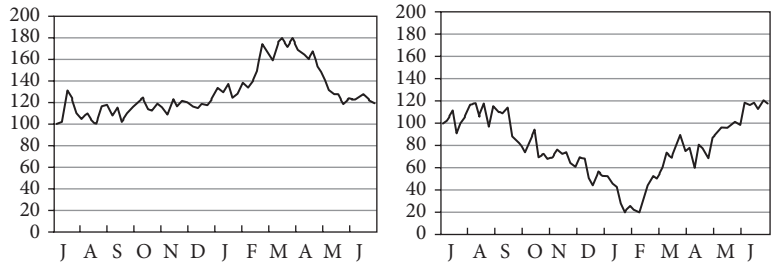


**Figure 19.4** High price is salient (left panel) and low price is salient (right panel). Starting and ending prices are the same in both cases. The average profits over the 12 months are 20% in both cases.

Source: Mussweiler, T. and K. Schneller (2003). "What goes up must come down—how charts influence decisions to buy and sell stocks," *The Journal of Behavioral Finance* 4(3): 121–30. Reprinted by permission of Taylor & Francis Ltd., http://www.tandfonline.com

panel has a salient low price. Otherwise the starting and the ending prices as well as the average profits corresponding to both panels are identical. Subjects were asked to state price expectations of the stock in each case and also asked if they would invest in the asset.

The anchoring heuristic suggests that subjects who observe the high salient price will have relatively higher price expectations and are more likely to invest as compared to subjects who are shown the low salient price. These predictions are confirmed. The predicted mean prices when the high and low price, respectively, is salient are 132 and 112; the difference is statistically significant.

## 19.6.2 *Legal and regulatory implications of anchoring*

Experienced individuals also exhibit the anchoring effect. The subjects in Englich and Mussweiler (2001) are experienced trial judges with an average experience of 15 years. In one experiment, they were given the case of a fictitious shoplifter who is accused of stealing from a supermarket for the 12th time. Detailed case material was compiled and it included opinions from a psycho-legal expert, and testimonies by the defendant and a witness. A pre-test with experienced legal professions showed that the case material was complete and realistic. In the pre-test, the mean sentence was 5.62 months with a standard deviation of 2.57.

Subjects in the experiment were asked to determine an appropriate sentence. In two different treatments, the case material showed that the prosecutor had asked for a sentence of, respectively, 9 months (high anchor) and 3 months (low anchor). Subjects in the high anchor treatment chose a sentence of 8 months and those in the low anchor treatment chose a sentence of 5 months. The *anchoring effect* was $\frac{8-5}{9-3} \times 100 = 50.0\%$. Crucially, the authors find that experience does not mitigate the anchoring effect. Legal professionals and experienced judges exhibit a similar bias.

Anchoring has interesting regulatory implications. Consider, for instance, the problem of caps on damages in litigation. Saks et al. (1997) considered two treatments. In the first treatment, damages were capped at the relatively high level of $250,000 for trial awards for minor injuries. In the second treatment, there were no caps. They found that the average trial award was $3895 in the uncapped treatment and $15,718 in the capped treatment; anchoring at a high cap level significantly increased the average trial award.

Babcock and Pogarsky (1999) considered the rate of pretrial settlements with and without caps. They conducted experiments in which subjects in their role as either plaintiffs or defendants received case material on a personal injury lawsuit. They then had to negotiate a settlement with

the other party within 20 minutes. The default award made by a judge, should the bilateral negotiation fail, was capped at $250,000 in one treatment and left uncapped in another treatment. It was found that the rate of pretrial settlement was higher in the capped treatment.

Pogarsky and Babcock (2001) found that the rate of pretrial settlement is negatively influenced if the damages cap is set very high. The consider a cap of $1,000,000 in the capped treatment and reduced the severity of the plaintiff's injuries in the case material, relative to Babcock and Pogarsky (1999). This ensured that the cap appeared to any reasonable observer as too high. Unknown to the subjects, an actual trial court judge had opined a settlement of $325,000, based on the case material. The non-binding settlement cap reduced the settlement rate, as the authors had expected; 79% of the plaintiff–defendant pairs settled in the uncapped treatment, while 67% settled in the capped treatment. Thus, litigants appear to look ahead and anticipate the effect on the award of damages that is caused by high caps.

### 19.6.3 *Robustness of anchoring*

The anchoring phenomenon is remarkably robust to a wide range of domains, contexts, and frames. These include estimates of price (Mussweiler et al., 2000; Northcraft and Neale, 1987); the probability of a nuclear war (Plous, 1989); the evaluations of lotteries and gambles (Chapman and Johnson, 1994); issues of legal judgment (Chapman and Bornstein, 1996; Englich and Mussweiler, 2001); and first offers in price negotiation (Galinsky and Mussweiler, 2001). Anchoring has been used to explain a range of other judgment heuristics. These include the hindsight bias (Fischhoff, 1975); preference reversals (Lichtenstein and Slovic, 1971); and non-linear probability weighting (Tversky and Kahneman, 1974).

Anchoring is also robust to a wide variety of conditions. These include, clearly uninformative/random anchors (Mussweiler and Strack, 2000b; Tversky and Kahneman, 1974), or anchors known to be very extreme (Chapman and Johnson, 1994; Strack and Mussweiler, 1997). Anchoring appears not to be affected by forewarning the subjects or by the provision of incentives (Tversky and Kahneman, 1974; Wright and Anderson, 1989; Wilson et al., 1996; Chapman and Johnson, 2002). We have already noted that experienced subjects also exhibit the anchoring effect. In conjunction, these findings suggest that humans appear hardwired to use the anchoring heuristic.

### 19.6.4 *Explanations of anchoring*

In light of the robustness and reliability of the phenomenon, it is important to ask what causes anchoring. Tversky and Kahneman (1974) conjectured an explanation in terms of *insufficient adjustment* from an anchor to a proposed target. Quattrone (1982) and Quattrone et al. (1981) propose a bounded rationality explanation for insufficient adjustment. In this proposal, the individual begins with an *anchor,* $x_0$, about some target value, $x_T$, that is believed to belong to an interval $[a, b]$, where $a, b$ are real numbers. Suppose that $x_0 \notin [a, b]$. The individual then engages in a series of improvements of the anchor using a *satisficing heuristic* and stops whenever he hits one of the boundaries of the interval $[a, b]$. If the guess that $x_T \in [a, b]$ is objectively correct, the individual ends up at a distance $x - a$ or $b - x$ from the actual value, $x$. This is referred to as insufficient adjustment. However, if $x_0 \in [a, b]$, then no adjustment is needed.

Evidence for the insufficient adjustment hypothesis is given in LeBoeuf and Shafir (2006, 2009) who were inspired by the original experiments in the psychophysics of anchoring. In LeBoeuf and Shafir (2006), subjects were required to draw a line that was 3.5 inches (or 89 mm) long; this corresponds to $x_T$. They were given one of two anchors; see Figure 19.5. The short anchor, $x_0^s$, was a line starting from the bottom of the page; subjects were told that this line was shorter than 3.5

**Figure 19.5** Anchoring and insufficient adjustment.

Source: *Journal of Behavioral Decision Making* 19(4): 393–406. LeBoeuf, R. and Shafir, E., "The long and short of it: physical anchoring effects." Copyright © 2006 John Wiley & Sons, Ltd.

inches. Subjects were asked to extend it, so that, in their assessment, it was 3.5 inches (88.9 mm) in length. Another group of subjects were shown a line starting from the top of the page and told that it was longer than 3.5 inches; this is the long anchor, $x_0^l$. They had to shorten the long line to 3.5 inches. Subjects exposed to $x_0^s$ produced, on average, a line length of 61.2 mm with a standard deviation of 17 mm, and subjects exposed to $x_0^l$ produced a line that had an average length of 74.7 mm with a standard deviation of 15.4 mm.

A second explanation for anchoring is that anchors induce System 1 in the brain to search for anchor-consistent information. In this explanation, the anchor serves a *priming function*. Consider Tversky and Kahneman's (1974) experiments on anchoring described above in which subjects are asked for the number of African countries in the UN. Suppose that the wheel of fortune comes up with a high number, which is the higher anchor. A high anchor enables one to access information that is consistent with the number of countries in Africa being high. For instance, that Africa has a large size, or Africa has a large coastline.

It is crucial to appreciate that System 1 attempts to evaluate the given information by trying to prove that it is true (Snyder and Swan, 1978; Crocker, 1982; Trope and Bassok, 1982; Klayman and Ha, 1987). As Kahneman (2011) argues, System 1 tries its best to construct a world in which the anchor is the true number. System 2 is deliberative, yet it deliberates on the menu of memories chosen by System 1. Hence, a biased set of memories, evoked by priming, have a ripple effect on the decision made by the individual in the absolute judgment task.

When should we evoke the insufficient adjustment explanation and when should we evoke the priming explanation? Epley and Gilovich (2001) give one possible resolution of this problem. They distinguish between two kinds of anchors. Those that are self-generated and those that are provided by the experimenter. They argue that self-generated anchors are known to be wrong from the beginning, so the individual does not need to engage in retrieving anchor-consistent information to judge its correctness. Rather, the problem is to adjust the initial anchor towards the target, as explained above in the case of insufficient adjustment. However, anchors provided by the experimenter need to be evaluated for correctness by evoking anchor-consistent information.

Consider two examples of self-generated anchors from Epley and Gilovich (2001). (1) Subjects are asked to estimate the freezing point of vodka. In thinking about the answer, participants typically start at 0°C, the freezing point of water, and adjust downwards, knowing that the freezing

point of alcohol is lower than water. (2) Which year was George Washington first elected president? In this case, most Americans seem to begin with the year of the declaration of independence, 1776, and adjust upwards from it.

Epley and Gilovich find that self-generated anchors trigger adjustment, while those provided by the experimenter do not. For instance, in the second question (election year of George Washington), 64% of the subjects were found to engage in anchoring and adjustment when the anchor was self-generated. However, when the anchor was provided by the experimenter in two other questions, mean length of a whale, and mean winter temperature in Antarctica, only 12% and 14%, respectively, engaged in anchoring and adjustment. This reasoning is confirmed in the experimental results of Epley and Gilovich (2006). Furthermore, they find that adjustment entails cognitive effort. When subjects are exposed to cognitive load, say, having to remember a set of numbers, or they are drunk, then they engage in insufficient adjustment.

## 19.7 Base rate neglect and conservatism

The use of Bayes' rule underpins all aspects of neoclassical economics. In this section, we examine the evidence on Bayes' rule. Suppose that we have $n$ mutually exclusive random variables $X_1, X_2, \ldots, X_n$ such that the sample space $X = X_1 \cup X_2 \cup \ldots \cup X_n$. Then Bayes' rule implies the following computation for any event $A \subseteq X$ such that $P(A) > 0$,

$$P(X_i \mid A) = \frac{P(A \mid X_i)P(X_i)}{P(A)} = \frac{P(A \mid X_i)P(X_i)}{\sum_{j=1}^{n} P(A \mid X_j)P(X_j)}. \tag{19.8}$$

In (19.8), $P(X_i)$ and $P(A)$, $i = 1, \ldots, n$, are referred to as the *base rates*.

### 19.7.1 *Base rate neglect*

An important empirical finding is that base rates have very low salience in calculations of $P(X_i \mid A)$. Meehl and Rosen (1955) is an early study that found underweighting of base rates by clinical psychologists. Consider the following well-known *cab problem* introduced by Kahneman and Tversky (1972); modified versions of the problem are considered in Bar-Hillel (1980) and Tversky and Kahneman (1980).

There are only two cab companies in the city, Green and Blue; 85% of the cabs are Green. There was an accident last night. A witness comes forward to testify that the cab involved in the accident was Blue. In similar conditions, the reliability of the witness is 80%, i.e., the probability that the witness gets it wrong is 20%. What is the probability that the actual cab involved in the accident was Blue?

Let $X_1$ = Blue cab, $X_2$ = Green cab, and let the event, "identification by the witness that the cab is Blue" be $A$. Then, using (19.8),

$$P(X_1 \mid A) = \frac{P(A \mid X_1)P(X_1)}{P(A \mid X_1)P(X_1) + P(A \mid X_2)P(X_2)}. \tag{19.9}$$

Using $P(A \mid X_1) = 0.8$ (probability of a correct identification), $P(A \mid X_2) = 0.2$ (probability of an incorrect identification), $P(X_1) = 0.15$ (base rate), and $P(X_2) = 0.85$ (base rate) we get

$$P(X_1 \mid A) = 0.414.$$

By contrast, the median and the modal response was $P(X_1 \mid A) = 0.8$. Notice that 0.8 is also the probability with which the witness correctly identifies a Blue cab as the one involved in the accident. In particular, individuals appear to underweight base rates, which is not consistent with Bayes' rule. They ignore the fact that only 15% of the taxis are actually Blue. The problem of underweighting of base rates diminishes but does not disappear when greater salience is given to base rates in the design of experiments (Ajzen, 1977; Bar-Hillel, 1980).[14]

Consider an adaptation of the problem that Casscells et al. (1978) gave to students at Harvard Medical School. Suppose that a police Breathalyzer test discovers false drunkenness in 5% of the cases when the driver is sober. However, the Breathalyzers are always able to detect a truly drunk person with complete certainty. In the general population, a fraction 1/1000 of drivers engage in driving while drunk. Suppose that a driver is checked at random, and takes a Breathalyzer test, which shows that the driver is drunk. How high is the probability he or she really is drunk? The modal response was 95%, the mean response was 56% and a sixth of the students gave the correct answer, which is approximately 2%. Thus, even very educated and clever subjects underweight the base rate in relatively simple problems.

In some cases, underweighting of base rates does not take place. Consider, for instance, the following variant of the cab problem from Kahneman (2011). "There are only two cab companies in the city, Green and Blue, who own an equal number of cabs. 85% of the cabs involved in accidents are Green." The rest of this problem is identical to the earlier cab problem following the instruction: "There was an accident last night" . . . This, and the earlier cab problem, are mathematically equivalent. Yet, one finds that subjects give attention to the base rate in the second problem. The median response is that the cab involved in the accident was blue is 0.60. Indeed this appears as an average of the correct answer, 0.41, and the reliability of the witness, 0.80.

What accounts for the relative salience of base rates in this case? The answer depends on the distinction made by Ajzen (1977) between *causal* and *incidental* base rates. A causal base rate conveys information on some causal factors relevant to the problem, while incidental base rates do not. For instance, in the original cab problem, the base rates are incidental. However, in the modified cab problem, the base rates are causal because they suggest to the decision maker that drivers of Green cabs are particularly reckless. This is because despite identical numbers of Blue and Green cabs, Green cabs are involved in 85% of the accidents. Hence, the individual assigns a lower probability to Blue cabs being involved in the accident, which curbs their tendency to underweight the base rate of Blue cabs.

Bar-Hillel (1980) showed that subjects may combine information on base rates with other non-causal information to mitigate base rate neglect. For instance, in one modification of the cab problem, subjects were told that the hit-and-run cab was equipped with an intercom and intercoms are installed on 80% of the Green cabs and 20% of the Blue cabs. The mathematical structure of this problem is identical to the original cab problem, yet the median response was 0.48, not far from the Bayesian answer of 0.41.

A variety of factors have been identified that reduce base rate neglect. Individuals appear to pay more attention to base rate information, when they are instructed to think like statisticians rather than clinical psychologists (Schwarz, Strack, Hilton, and Naderer, 1991). Individuals with higher measured intelligence also underweight base rates less (Stanovich and West, 2000). Girotto and

[14] For more evidence on the neglect of base rates, see Grether (1980) who found that likelihoods are more salient than base rates, although the latter were not altogether ignored. See also Grether (1992) for an extension of his earlier work.

Gonzalez (2001) find that if the attention of subjects is drawn to the computations required in a Bayes' rule calculation, then base rate neglect falls. Sloman et al. (2003) find that presenting the information in a diagram that highlights set inclusion also reduces base rate neglect.

Empirically it has been observed that presenting the information in a frequency format rather than a probability format reduces base rate neglect (Brase, 2002b; Cosmides and Tooby, 1996; Gigerenzer and Hoffrage, 1995; Pinker, 1997). Consider the following problem of *base rate neglect* used originally in Eddy (1982) and then in the experiments of Gigerenzer and Hoffrage (1995). The problem is expressed in probability terms as follows.

The probability of breast cancer is 1% for a woman at age forty who participates in routine screening [base-rate]. If a woman has breast cancer, the probability is 80% that she will get a positive mammography [hit-rate]. If a woman does not have breast cancer, the probability is 9.6% that she will also get a positive mammography [false-alarm rate]. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

The Bayesian solution is readily found. Let $X$ be the event "positive mammography" and $Y$ the event "breast cancer" and let $\sim X, \sim Y$ be negations of these events. Then

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X \mid Y)P(Y) + P(X \mid \sim Y)P(\sim Y)} = \frac{0.008}{0.008 + 0.096(0.99)} = 0.0776.$$

Thus, the Bayesian response is that conditional on getting a positive mammography, the probability that the individual has breast cancer is 7.76%. In Eddy's (1982) original study, less than 5% of the respondents gave answers that conformed to the Bayesian solution. Gigerenzer and Hoffrage (1995) tested the same problem in the probability format and the natural frequency format. In the probability format they found that the choices of 16% of the respondents conformed to the Bayesian solution, but in the natural frequency format, 46% of the responses conformed to Bayes' rule.

A sample of the studies that use the frequency format is shown in Table 19.4, drawn from Barbey and Sloman (2007). The second column shows that when the problem is presented in a probabil-

**Table 19.4** Percentage of responses consistent with Bayes' rule in different empirical studies. Sample sizes in parenthesis.

| Study | Information format and judgment domain | |
|---|---|---|
| | Probability | Frequency |
| Cascells et al. (1978) | 18(60) | – |
| Cosmides & Tooby (1996; Exp. 2) | 12 (25) | 72 (25) |
| Eddy (1982) | 5 (100) | – |
| Evans et al. (2000; Expl) | 24 (42) | 35 (43) |
| Gigerenzer (1996b) | 10 (48) | 46 (48) |
| Gigerenzer and Hoffrage (1995) | 16 (30) | 46 (30) |
| Macchi (2000) | 6 (30) | 40 (30) |
| Sloman et al. (2003; Exp 1) | 20 (25) | 51 (45) |
| Sloman et al. (2003; Exp lb) | — | 31 (48) |

Source: Barbey and Sloman, 2007.

ity format, the percentage of subjects conforming to Bayes' rule is relatively small. However, there is much greater conformity with Bayes' rule when the information is presented in a frequency format (third column). The Cosmides and Tooby (1996) result is an outlier that Sloman et al. (2003) and Evans et al. (2000) could not replicate. Despite the rhetoric, in the presence of a frequency format, less than half of the subjects conform to Bayes' rule in this cross section of studies.

It is also likely that other factors could facilitate conformity with Bayesian reasoning when the natural frequency format is given (Cosmides and Tooby, 1996). These factors include, in particular, monetary incentives and controlling for the ability levels of students, e.g., hiring subject pools at top universities. It is argued that these factors lead to statistically significant differences in the facilitation role of natural frequencies relative to the probability format (Brase et al., 2006). But this also casts doubt on the evolutionary explanation, which lies at the heart of the facilitating role of natural frequency formats. At best, these results are suggestive of the relation between cognitive effort and cognitive ability with the facilitation role of natural frequencies in improving conformity with Bayes' rule.

When data is presented in natural frequencies, but the set inclusion relation is not made obvious (although it can be easily computed from the data), then the facilitation role of natural frequencies drops significantly (Girotto and Gonzalez, 2001). Furthermore, Brase (2002a) finds that the probability and the natural frequency format are both perceived to be equally very clear unless the probabilities are of a very low magnitude. This is difficult to reconcile with the claimed facilitating role of the natural frequency format.

Based on the evidence, the case for the natural frequency format in reversing base rate neglect appears overstated.

Why might problems expressed in terms of natural frequency facilitate Bayesian reasoning? Some of the competing theories considered by Barbey and Sloman (2007) are as follows. In the *mind as a Swiss army knife view* (Cosmides and Tooby, 1996; Gigerenzer and Selten, 2001), the brain comprises several modules. Each module is responsible for a specific set of functions and is not under the voluntary control of the individual. In this view, there is a specific module, with evolutionary origins, designed to process natural frequencies, but not probabilities. One natural advantage of the frequency format is that it preserves the sample size and makes transparent the subset relation between two sets (e.g., 1 out of every 100 women have breast cancer).

A second view is that the brain has a *natural frequency algorithm* (Gigerenzer and Hoffrage, 1995). This view makes no assumptions about the modularity of brain architecture. However, for all practical purposes, this is fairly similar to the first view. Barbey and Sloman's (2007) preferred explanation is based on a System 1, System 2 distinction. In this view, the frequency format is more consistent with the natural inputs required by System 2 to process information. As the peer commentary in the same issue of *Behavioral and Brain Sciences* makes clear, however, there is no consensus among psychologists on this issue.

### 19.7.2 *Conservatism*

*Conservatism* refers to the underweighting of the likelihood of a sample. We illustrate this using a problem in Camerer (1995), which is originally due to Edwards (1968). Suppose that there are two urns $A$ and $B$, each of which contains only red and blue balls. Urn A has 7 red and 3 blue balls and urn B has 3 red and 7 blue balls. It is common knowledge that nature is equally likely to choose one of the two urns, so $P(A) = 0.5$ and $P(B) = 0.5$. A sample of 12 balls is drawn *with replacement* from one of the urns, but subjects do not know which one.

Suppose that the number of red balls in a sample of size 12 is $x \in \{0, 1, \ldots, 12\}$. Consider the sample, 8 red and 4 blue balls, which does not exactly represent the proportions in any of the two urns. Hence, we cannot use the representativeness heuristic. Let $p_A$ and $p_B$ be the respective probabilities of drawing a red ball from urns A and B. Analogous to (19.2), the conditional probability of drawing $x$ red balls in a sample of 12 balls from urn $i = A, B$ is

$$P(x \mid i) = \left( \begin{array}{c} 12 \\ x \end{array} \right) p_i^x (1 - p_i)^{12-x}; i = A, B \text{ and } x \in \{0, 1, \ldots, 12\}. \tag{19.10}$$

The Bayesian posterior belief that the sample came from urn $A$ is given by

$$P(A \mid x = 8) = \frac{P(x = 8 \mid A)\, P(A)}{P(x = 8 \mid A)\, P(A) + P(x = 8 \mid B)\, P(B)}. \tag{19.11}$$

We already know that $P(A) = 0.5, P(B) = 0.5$. Since $p_A = 0.7, p_B = 0.3$, we get from (19.10) that $P(x = 8 \mid A) = 0.231$, $P(x = 8 \mid B) = 0.008$. Substituting these numbers in (19.11), we get that $P(A \mid x = 8) = 0.967$. However, the typical response reported in Camerer (1995) is 0.7 to 0.8. Hence, subjects in experiments appear too conservative about sample likelihoods.

### 19.7.3 *A reconciliation of conservatism and base rate neglect*

It appears contradictory that in some cases the base rate is underweighted (as in the cab problem), while in others the sample information is underweighted (conservatism). We need a model that identifies circumstances when we are confident (underweighting base rates), and underconfident (conservatism). Griffin and Tversky (1992) show that these two phenomena can be reconciled by distinguishing between the *strength* and the *weight* of evidence. Consider two illustrative examples of these concepts.

1. Suppose that one is trying to evaluate a reference letter for a graduate student, written by an academic referee. The strength of the letter is the information that it provides about the student, for instance, how warm and positive the letter is, while the weight is the credibility/standing of the referee.
2. Suppose that we toss a coin $n$ times to determine if it is biased. The strength of the evidence is the number of times it shows up as heads while the weight is the sample size.

While it is not always possible to neatly separate the strength and weight of evidence in every case, in many cases one may be able to do so. Griffin and Tversky (1992) argue that people give undue importance to strength as compared to the weight of the evidence, relative to what would be prescribed by standard statistical techniques. In particular, people begin by giving undue importance to the strength of the evidence and then adjusting insufficiently for the weight of the evidence. For instance, in the first example, we often see how warm the reference letter is and then adjust for the credibility of the writer. We feel underconfident when there is a moderately positive letter from a very credible referee. An overconfident judgment arises when strength is high relative to weight. Conversely, an underconfident judgment gives relatively more importance to the weight, particularly when the strength is low.

Consider the following experiment that Griffin and Tversky conducted with 35 students. A coin is tossed $n$ times and the number of heads ($H$) and tails ($T$) is noted. Subjects are told that the bias of the coin is $3/5 = 0.6$, but they are not told the direction of the bias. Hence, subjects believe

that either $P(H) = 3/5$, or $P(T) = 3/5$. Denote the events, "head-biased coin" and "tail-biased coin," respectively, by $H_b$ and $T_b$. Prior probabilities of these mutually exclusive and exhaustive events are 50 : 50,

$$P(H_b) = P(T_b) = 0.5. \tag{19.12}$$

In the experiments, the sample size, $n$, varies from 3 to 33, while the number of heads varies from 2 to 19; see Table 19.5 for the results. Consider a sample of size $n$ in which there are $h_n$ heads and $n - h_n$ tails. Denote this sample by $h_n$. The posterior distribution $P(H_b \mid h_n)$ that the coin is head-biased, conditional on the sample $h_n$, is given by

$$P(H_b \mid h_n) = \frac{P(h_n \mid H_b)P(H_b)}{P(h_n \mid H_b)P(H_b) + P(h_n \mid T_b)P(T_b)}. \tag{19.13}$$

Using the formula for the binomial distribution, we can compute $P(h_n \mid H_b)$ and $P(h_n \mid T_b)$ and rewrite (19.13) as

$$P(H_b \mid h_n) = \frac{\binom{n}{h_n}\left(\frac{3}{5}\right)^{h_n}\left(\frac{2}{5}\right)^{n-h_n}}{\binom{n}{h_n}\left(\frac{3}{5}\right)^{h_n}\left(\frac{2}{5}\right)^{n-h_n} + \binom{n}{h_n}\left(\frac{2}{5}\right)^{h_n}\left(\frac{3}{5}\right)^{n-h_n}}. \tag{19.14}$$

The first three columns in Table 19.5 give the experimental data, and the final column gives the computation of posterior probabilities, based on (19.14).

In an analogous manner, we get

$$P(T_b \mid h_n) = \frac{P(h_n \mid T_b)P(T_b)}{P(h_n \mid T_b)P(T_b) + P(h_n \mid H_b)P(H_b)}. \tag{19.15}$$

**Table 19.5** Probability that the coin is head-biased, conditional on the observations.

| $n$ | $h_n$ | $n - h_n$ | $P(H_b \mid h_n)$ |
|---|---|---|---|
| 3 | 2 | 1 | 0.60 |
| 3 | 3 | 0 | 0.77 |
| 5 | 3 | 2 | 0.60 |
| 5 | 4 | 1 | 0.77 |
| 5 | 5 | 0 | 0.88 |
| 9 | 5 | 4 | 0.60 |
| 9 | 6 | 3 | 0.77 |
| 9 | 7 | 2 | 0.88 |
| 17 | 9 | 8 | 0.60 |
| 17 | 10 | 7 | 0.77 |
| 17 | 11 | 6 | 0.88 |
| 33 | 19 | 14 | 0.88 |

Source: Griffin and Tversky (1992).

Dividing (19.13) by (19.15) and using (19.12), we get

$$\frac{P(H_b \mid h_n)}{P(T_b \mid h_n)} = \frac{P(h_n \mid H_b)}{P(h_n \mid T_b)}. \tag{19.16}$$

Recall that the sample $h_n$ denotes $h$ heads out of a sample of $n$ iid draws from either a heads-biased coin ($P(H) = 0.6$) or a tail-biased coin ($P(T) = 0.6$). Since the successive tosses of the coin are independent, the likelihoods of the sample under, respectively, the heads-biased coin and the tail-biased coin, are given by

$$P(h_n \mid H_b) = (0.6)^{h_n}\,(0.4)^{n-h_n}\,; P(h_n \mid T_b) = (0.4)^{h_n}\,(0.6)^{n-h_n}\,. \tag{19.17}$$

Substituting (19.17) in (19.16) and denoting by $t_n = n - h_n$, the number of tails in the sample, we get

$$\frac{P(H_b \mid h_n)}{P(T_b \mid h_n)} = \left(\frac{0.6}{0.4}\right)^{h_n - t_n}.$$

Taking logs on both sides and multiplying and dividing the right hand side by $n$, we get the following form for Bayes' rule.

$$\log\left(\frac{P(H_b \mid h_n)}{P(T_b \mid h_n)}\right) = n\left(\frac{h_n - t_n}{n}\right)c;\ \text{where } c = \log\left(\frac{0.6}{0.4}\right),\ \text{a constant.} \tag{19.18}$$

We can now use (19.18) to further clarify the notions of strength and weight, using terms on the right hand side.

1. The first term, $n$, is the *weight* of the evidence.
2. The second term, the difference between the proportions of heads and tails in the sample, is the *strength* of the evidence for $H_b$ and against $T_b$; the experiment only considers the cases such that $h_n > t_n$.
3. The third term, $c$, is the *discriminability* of the evidence.

Taking logs on both sides of (19.18) we get

$$\log\log\left(\frac{P(H_b \mid h_n)}{1 - P(H_b \mid h_n)}\right) = \log n + \log\left(\frac{h_n - t_n}{n}\right) + \log c. \tag{19.19}$$

Griffin and Tversky (1992) estimate a regression of the following form:

$$y = \beta_0 + \beta_1 \log n + \beta_2 \log\left(\frac{h_n - t_n}{n}\right) + \beta_3 \log c + u \tag{19.20}$$

where $y = \log\log\left(\frac{P(H_b|h_n)}{P(T_b|h_n)}\right)$, $u$ is a random error, and $\beta_0, \beta_1, \beta_2, \beta_3$ are the regression coefficients. To be consistent with Bayes' rule, (19.18), the predicted regression coefficients are: $\beta_1 = \beta_2 = 1$. However, for 30 out of 35 subjects, $\beta_2 > \beta_1$ and the difference is statistically significant. Across subjects, the median ratio $\frac{\beta_2}{\beta_1} = 2.2$, i.e., subjects weighted the strength more than twice as much as the weight.

Since the conditional probability function is also a probability function, $P(H_b \mid h_n) + P(T_b \mid h_n) = 1$. Fix $P(H_b \mid h_n)$ at some number in the open interval $(0,1)$, so that the LHS of (19.19) equals a constant, $\bar{y}$. An example is the value of $P(H_b \mid h_n) = 0.77$ for $n = 3$ and 9 (see Table 19.5). Thus, (19.19) can be written as

$$\log\left(\frac{h_n - t_n}{n}\right) = (\bar{y} - \log c) - \log n. \tag{19.21}$$

We now ask, what combinations of strength, $\frac{h_n - t_n}{n}$, and weight, $n$, satisfy equation (19.21), given that $\bar{y} - \log c$ is a fixed number? A curve that expresses such combinations is known as an *equal-support line*. In $\left(\log n, \log\left(\frac{h_n - t_n}{n}\right)\right)$ space, the equal-support line based on (19.21) has a slope $-1$ and an intercept that is increasing in $P(H_b \mid h_n)$ (because $\bar{y}$ is increasing in $P(H_b \mid h_n)$). So, for instance, if we fixed $P(H_b \mid h_n) = 0.77$, then, using Table 19.5, (19.21) is satisfied for the following two (strength, weight) pairs: $\left(\frac{h_n - t_n}{n}, n\right) = (1, 3), \left(\frac{1}{3}, 9\right)$.

In Figure 19.6, we plot two equal-support lines that are consistent with Bayesian updating (i.e., $\beta_1 = \beta_2 = 1$). These are shown as the two dotted lines. The upper one corresponds to a higher value of $\bar{y} = y_H$, relative to the lower one, $\bar{y} = y_L < y_H$.

We can also derive the equal support lines from the estimated regression. Rewrite (19.20) as follows.

$$\log\left(\frac{h_n - t_n}{n}\right) = \frac{1}{\beta_2}\left(y - \beta_0 - \beta_3 \log c - u\right) - \frac{\beta_1}{\beta_2}\log n. \tag{19.22}$$

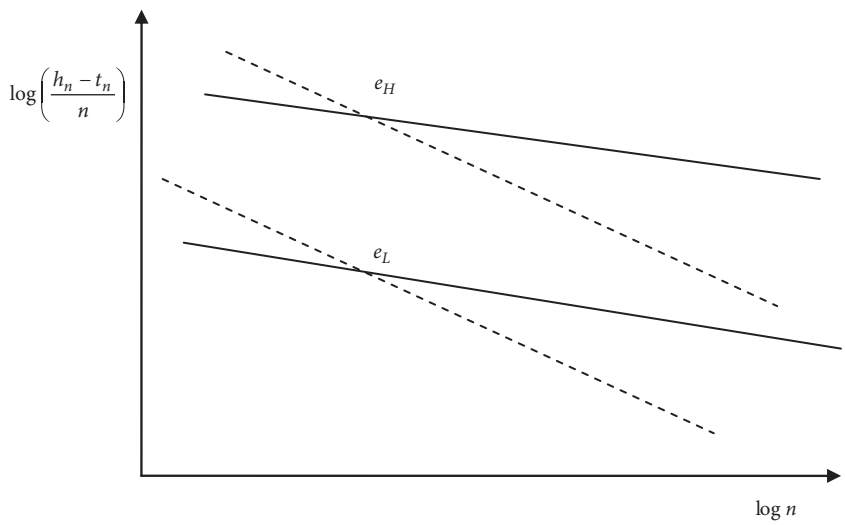


**Figure 19.6** Determining the relative importance of the strength and the weight of evidence.

Source: Reprinted from *Cognitive Psychology* 24(3): 411–35. Griffin, D. and Tversky, A. (1992). "The weighing of evidence and the determinants of confidence." Copyright © 1992, with permission from Elsevier.

Since, empirically, subjects place greater salience on strength, relative to weight, $\frac{\beta_1}{\beta_2} < 1$, so the slope of the estimated equal-support line is flatter. Two such estimated equal-support lines are drawn for $y = y_H$ and $y = y_L$; these are shown as the two continuous lines.

Consider the intersection points $e_H$ and $e_L$, where the Bayesian line intersects the estimated line. To the left of these intersection points (for any level of weight, $\log n$) the estimated strength is lower than the correct statistical (or Bayesian) strength. Individuals are said to be *underconfident* in this region. Conversely, to the right of these intersection points, estimated strength is greater than the Bayesian strength. Individuals are said to be *overconfident* in this region. Conservatism is a sort of *underconfidence* in the sample likelihood. On the other hand, base rate neglect arises when individuals are *overconfident* of the strength of the evidence. Hence, points such as $e_H$, $e_L$ divide the space into regions where conservatism and base rate neglect arise naturally.

Consider an individual engaged in the stock market. When he hears isolated bits of news, he views the news as having low strength, so he does not find this information very useful (although it might have high weight; for instance, the news might be a report of an assessment by a leading expert). On the other hand, when the individual hears about a small sequence of good news about a company selling some stock, he might assign much higher strength to it as compared to a Bayesian. The individual might overreact to a sequence of good news based on an overweighted strength. However, the news might have little weight in predicting the future performance of the company.

## 19.8 Hindsight bias

Suppose that an individual is asked at time $t$ to predict the outcome of an uncertain event, $X$, that will take place at some time $t + j$, $j > 0$. Denote the information set at time $t$ by $I_t$. Then the conditional prediction, denoted by $E[X \mid I_t]$, reflects the individual's *predictive judgment*. Suppose that at time $t + j$, the realization of the random variable is $x$. Posterior to the observation of $x$, the individual is asked about the remembered prediction or the *postdictive judgment*. This is given by $E\left[E[X \mid I_t] \mid I_{t+j}\right]$, where $I_{t+j}$ is the date $t + j$ information set, and $I_t \subseteq I_{t+j}$. In neoclassical economics, the typical assumption is $E[X \mid I_t] = E\left[E[X \mid I_t] \mid I_{t+j}\right]$. An individual is said to suffer from *hindsight bias*, or *creeping determinism* (Fischhoff, 1975), if the predictive and postdictive judgments differ, i.e.,

$$E[X \mid I_t] \neq E\left[E[X \mid I_t] \mid I_{t+j}\right], \tag{19.23}$$

and the postdictive judgment is biased in favour of the actual outcome. For instance, Camerer et al. (1989) incorporate hindsight bias by the following simple rule

$$E\left[E[X \mid I_t] \mid I_{t+j}\right] = \alpha x + (1 - \alpha)E[X \mid I_t]; \alpha \in [0, 1]. \tag{19.24}$$

The individual has no hindsight bias if $\alpha = 0$ and is fully-biased if $\alpha = 1$. Hindsight bias is sometimes termed as the "I knew it all along" effect; this is particularly obvious when $\alpha = 1$. Hindsight bias is quite distinct from learning. In learning, one learns to make better predictions of future events. However, under hindsight bias, the relevant issue is remembered past predictions, i.e., postdictive judgments.

## 19.8.1 *Empirical evidence on the hindsight bias*

In his pioneering work, Fischhoff (1975) ran six experiments in a between-subjects design; each experiment has a Before and After group. In each experiment, the Before group read a brief 150 word account of a historical or clinical event and then had to rate the probabilities of four possible mutually exclusive outcomes of which only one outcome actually occurred; this captures predictive judgment. The After group then read the same 150 word account, and was told that one of the four outcomes had occurred; all four outcomes were used in this stage, but each experimental group was given only one outcome and told that this was the real outcome. Based on this information, the After group was required to make a postdictive judgment.

Each of the six experiments used four events, A, B, C, and D, to construct the 150 word accounts. The events were chosen so that subjects were unlikely to know the actual outcome. For instance, event A used a historical account that was based on the British–Gurkha struggle. It was presented as follows (p. 305):

> For some years after the arrival of Hastings as governor-general of India, the consolidation of British power involved serious war. The first of these wars took place on the northern frontier of Bengal where the British were faced by the plundering raids of the Gurkhas of Nepal. Attempts had been made to stop the raids by an exchange of lands, but the Gurkhas would not give up their claims to country under British control, and Hastings decided to deal with them once and for all. The campaign began in November 1814. It was not glorious. The Gurkhas were only some 12000 strong; but they were brave fighters, fighting in territory well-suited to their raiding tactics. The older British commanders were used to war in the plains where the enemy ran away from a resolute attack. In the mountains of Nepal it was not easy even to find the enemy. The troops and transport animals suffered from the extremes of heat and cold, and the officers learned caution only after sharp reverses. Major-General Sir D Octerlony was the one commander to escape from these minor defeats.

The possible outcomes offered were: (1) British victory, (2) Gurkha victory, (3) military stalemate with no peace settlement, and (4) military stalemate with a peace settlement. The factually correct outcome was (3) but the After subjects were informed of one of the factual outcomes (1)–(4) as the correct one.

The average probabilities for each of the outcomes, 1–4, in the four events, A–D, are shown in Table 19.6. For each event, the first row gives the predictive judgment of the Before group. The next four rows give the postdictive judgments of the After groups, for each of the outcomes, 1–4, when told that outcomes 1, 2, 3, and 4, respectively, were the real outcome.[15] Hindsight bias arises if the before probabilities (first row) and the after probabilities (along the diagonal of the matrix) differ and, in particular, the after probabilities are higher. This was true in all cases, and in 22/24 of the cases, the after–before difference was statistically significant at the 2.5% level. The mean difference between the after and before probabilities was 10.8%. At the level of the individual subjects, 70% of the After subjects assigned higher probabilities relative to the mean before probabilities.

[15] Just to make sure that there is no confusion about Table 19.6, consider Event A (British–Gurkha struggle). On average, 33.8% of the before group assigns outcome 1 as the most likely outcome (predictive judgment). However, 57.2% of the after group that has been told that the outcome was 1, believes that the before group must have chosen outcome 1 as the most likely outcome (postdictive judgment); only 14.3% in the same group believe that the before group chose outcome 2. Similarly, 21.3% of the before group chooses outcome 2. However, 38.4% of the after group that was given outcome 2 as the actual outcome, believed that outcome 2 was chosen by the before group (postdictive judgment).

**Table 19.6** Mean probabilities assigned to each of four outcomes, 1–4, in the four events, A–D. The actual outcomes for events A, B, C, and D were, respectively, 1, 1, 4, and 2.

| Experimental group | n | Outcome provided | Outcome evaluated | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 |
| Event A: British–Gurka struggle (English-speaking subjects) | | | | | | |
| Before | 20 | None | 33.8 | 21.3 | 32.3 | 12.3 |
| After | 20 | 1 | 57.2 | 14.3 | 15.3 | 13.4 |
| | 20 | 2 | 30.3 | 38.4 | 20.4 | 10.5 |
| | 20 | 3 | 25.7 | 17.0 | 48.0 | 9.9 |
| | 20 | 4 | 33.0 | 15.8 | 24.3 | 27.0 |
| Event B: Near-riot in Atlanta (subjects with knowledge of statistics) | | | | | | |
| Before | 20 | None | 11.2 | 30.8 | 43.8 | 14.2 |
| After | 20 | 11 | 30.6 | 25.8 | 23.3 | 20.3 |
| | 20 | 2 | 5.5 | 51.8 | 24.3 | 18.5 |
| | 20 | 3 | 2.9 | 23.9 | 50.8 | 21.4 |
| | 20 | 4 | 16.7 | 31.9 | 23.4 | 27.9 |
| Event C: Mrs Dewar in therapy | | | | | | |
| Before | 19 | None | 26.6 | 15.8 | 23.4 | 34.4 |
| After | 13 | 11 | 43.1 | 13.9 | 17.3 | 25.8 |
| | 17 | 2 | 26.5 | 23.2 | 13.4 | 36.9 |
| | 16 | 3 | 30.6 | 14.1 | 34.1 | 21.3 |
| | 17 | 4 | 21.2 | 10.2 | 22.6 | 46.1 |
| Event D: George in therapy | | | | | | |
| Before | 17 | None | 27.4 | 26.9 | 39.4 | 6 3 |
| After | 18 | 11 | 33.6 | 20.8 | 37.8 | 8.0 |
| | 18 | 2 | 22.4 | 41.8 | 28.9 | 7.1 |
| | 20 | 3 | 20.5 | 22.3 | 50.0 | 7.3 |
| | 17 | 4 | 30.6 | 19.5 | 37.7 | 12.3 |

Source: Fischhoff (1975).

In order to address potential concerns about the between-subjects design of the earlier study, Fischhoff and Beyth (1975) used a within-subjects design. They asked Israeli subjects to make predictions about a set of 15 possible outcomes on the eve of President Nixon's visit to China and the USSR. Prior to the visit, the subjects filled in a prediction memory questionnaire. Following the actual visit of the President, subjects were asked to make postdictive judgments. Significant hindsight bias was discovered in the postdictive judgments.

In the context of the 9/11 tragedy in the US, Kahneman (2011) cites public reaction when it was revealed that on July 2001 the CIA had come across information to suggest that al-Qaeda might be planning an attack on the US. In this case, the CIA did not pass on the information to the President but did pass it on to the National Security Advisor. A former executive editor of the *Washington Post* wrote (p. 294): "It seems elementary to me that if you've got the story that is going to dominate history you might as well go right up to the President." However, in July 2001 nobody knew that this was going to make history.

A similar case of 20–20 judgment in hindsight is the proliferation of claims by experts following the 2008 financial crises that they had predicted the crash. From time to time, experts predict that a current economic situation may lead to a recession. However, in 2006/7 there was no widespread agreement among experts about the scale of the impending crash or certainty about the failure of Lehman Brothers. Kahneman (2011) infers from this that the crash was not *knowable*. Yet the postdictive judgment of many experts is that they knew it all along; this is an illustration of the *illusion of understanding*.

The hindsight bias has been found for all age groups from 3 to 95 (Bernstein et al., 2011). It has been found in many different contexts such as medical diagnosis (Arkes et al., 1988), accounting decisions (Anderson et al., 1993), football (Roese and Maniar, 1997), legal liability (Hastie et al., 1999; Goodwill et al., 2010), terrorist attacks (Fischhoff et al., 2005), the verdict in the O.J. Simpson criminal trial (Bryant and Brockway, 1997), and labor disputes involving firefighters (Pennington, 1981).

For attempts to debias subjects who suffer from hindsight bias, see Arkes (1991), Fischhoff (1982), Larrick (2004), and Harley (2007). The meta–study by Guilbault et al. (2004) suggests that expertise does not shield decision makers from the hindsight bias. Expertise is likely to mitigate a bias if experts receive continual feedback (weather forecasters, chess players, insurance firms) as compared to cases where less frequent feedback is available, or it is ambiguous (judges, stockbrokers, political experts); see, for instance, Kahneman and Klein (2009), McKenzie et al. (2008), and Tetlock (2006).

### 19.8.2 *Hindsight bias in legal situations*

The issue of hindsight bias is particularly important in legal situations. Consider a defendant who can take several possible actions; each action leads to a possible harm for others with some probability. Suppose that the chosen action of the defendant did indeed cause harm to a plaintiff, who then sues the defendant. A judge or a jury must now form a postdictive judgment of the reasonableness of the action taken by the defendant. In the presence of hindsight bias, the postdictive judgment is biased, so the defendant is more likely to be found guilty. Thus, liability is likely to be assigned in a biased manner, even if the defendant took the most reasonable action.

If potential defendants are forward looking, then anticipating potential future liability, they may take excessive precautions and engage in excessively risk averse behavior. For instance, doctors may prescribe more tests than they otherwise would have, or avoid minor surgical procedures, or be willing to pay for even more comprehensive and expensive insurance, which feeds into higher treatment costs.

Rachlinski (1998) argues that the legal system recognizes well the presence of hindsight bias and takes reasonable steps to mitigate it, as far as possible. For instance, in medical malpractice lawsuits, reliable ex-ante judgment of reasonable care is available; in this case, courts simply use this judgment as its legal postdictive judgment. Corporate law protects directors of corporations from liability for negligent business decisions. This eliminates the possibility of hindsight bias in legal judgment in this domain, enabling corporations to operate in an unfettered environment; this has its pros (fewer restrictions on business investments) and cons (socially irresponsible but privately optimal decisions). However, this does not apply to individuals in their private capacity, hence, the law views the negligence of private individuals as a form of strict liability; in this case, the outcome is third-best.[16]

[16] Our usage follows the standard sense of the term in economics: The first best obtains if no harm is done, the second best occurs if a judgment can be passed in the absence of hindsight bias and the third best occurs if judgment is passed with hindsight bias.

In Kamin and Rachlinski (1995), subjects make predictive and postdictive judgments about a municipality's actions to engage in flood precautions. The damages from the flood are set at $\$10^7$ and the cost of the precautions at $\$10^5$; subjects were also provided with meteorological and anecdotal data. One group of subjects, group A, were asked to predict the probability of a flood in the absence of precautions. They were told that if they find the probability greater than 10% they should decide to take the precaution. Another group of subjects, group B, was asked to make the postdictive judgment. They were told that the precautions had not been taken and a flood had occurred. They were told that if their postdictive judgment is that the probability of a flood was greater than 10% they should find the defendant guilty. In Group A, 24% of the subjects thought that the precautions were merited, while 57% in group B found the defendant guilty.

### 19.8.3 *What causes hindsight bias?*

What causes hindsight bias? Kahneman (2011, p. 202) opines: "A general limitation of the human mind is its imperfect ability to reconstruct past states of knowledge, or beliefs that have changed. Once you adopt a new view of the world (or of any part of it), you immediately lose much of your ability to recall what you used to believe before your mind changed." Other explanations are that (i) people prefer to see a stable and predictable world and the hindsight bias is simply a coping mechanism in the face of an unpredictable world, and (ii) by exaggerating their ability to predict, individuals attempt to signal greater competence to others (Rachlinski, 1998).

### 19.8.4 *Hindsight bias and the underestimation of financial volatility*

We now show that hindsight bias can lead decision makers to reduce volatility estimates in financial markets and lead to insufficient revision of incorrect models (Biais and Weber, 2009). Consider the two equations (19.23) and (19.24); to simplify notation, let $E[X \mid I_t] = \mu$ and let $E\left[E[X \mid I_t] \mid I_{t+j}\right] = \hat{\mu}$. Then we can write (19.24) as

$$\hat{\mu} = \alpha x + (1-\alpha)\mu. \tag{19.25}$$

Suppose that an investor at time $t$ forecasts the average returns at time $t+j$ on an asset, or a portfolio of assets. When making the prediction at time $t$, the investor believes that $X \sim N(\mu, \sigma_L^2)$ with probability $\lambda$ and $X \sim N(\mu, \sigma_H^2)$ with probability $1-\lambda$, where $\sigma_L^2 < \sigma_H^2$. Thus, the density of $X$ is given by

$$f(x \mid \sigma_i^2, \mu) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-1}{2\sigma_i^2}(x-\mu)^2\right), i = L, H. \tag{19.26}$$

At the end of time period $t+j$, after observing the realization, $x$, of the random variable, $X$, the decision maker forms posterior beliefs about which model (i.e., variance $\sigma_L^2$ or $\sigma_H^2$) is likely to have generated $x$. However, at the end of time $t+j$, on account of hindsight bias, the remembered mean is $\hat{\mu}$, given in (19.25). Thus, the relevant density, ex-post, is $f(x \mid \sigma_i^2, \hat{\mu})$; this may be used to determine, ex-post, which of the two competing models is more likely to hold. Denote by $\hat{P}(a \mid b)$ and $P(a \mid b)$, respectively, the probability of $a$ conditional on $b$ in the presence of hindsight bias ($\alpha > 0$) and in the absence of hindsight bias ($\alpha = 0$).

Using Bayes' rule, we have

$$\hat{P}(\sigma_L^2 \mid x) = \frac{f(x \mid \sigma_L^2, \hat{\mu})\lambda}{f(x \mid \sigma_L^2, \hat{\mu})\lambda + f(x \mid \sigma_H^2, \hat{\mu})(1-\lambda)}. \tag{19.27}$$

Using (19.26) to substitute the value of $f(x \mid \sigma_i^2, \hat{\mu})$, we can rewrite (19.27) as

$$\hat{P}(\sigma_L^2 \mid x) = \frac{1}{1 + \left(\frac{1-\lambda}{\lambda}\right)\frac{\sigma_L}{\sigma_H}\exp\left(\frac{-1}{2}\left(\frac{1}{\sigma_H^2} - \frac{1}{\sigma_L^2}\right)(x-\hat{\mu})^2\right)}. \tag{19.28}$$

Substituting $\hat{\mu}$ from (19.25) in (19.28), we get

$$\hat{P}(\sigma_L^2 \mid x) = \frac{1}{1 + \left(\frac{1-\lambda}{\lambda}\right)\frac{\sigma_L}{\sigma_H}\exp\left(\frac{-1}{2}\left(\frac{1}{\sigma_H^2} - \frac{1}{\sigma_L^2}\right)(1-\alpha)^2(x-\mu)^2\right)}. \tag{19.29}$$

In the absence of hindsight bias ($\alpha = 0$), we get the posterior probability, $P(\sigma_L^2 \mid x)$, as used in typical models in economics and finance

$$P(\sigma_L^2 \mid x) = \frac{1}{1 + \left(\frac{1-\lambda}{\lambda}\right)\frac{\sigma_L}{\sigma_H}\exp\left(\frac{-1}{2}\left(\frac{1}{\sigma_H^2} - \frac{1}{\sigma_L^2}\right)(x-\mu)^2\right)}. \tag{19.30}$$

Comparing (19.29) and (19.30), we derive the main implications of hindsight bias in this model. From (19.29), since $\frac{-1}{2}\left(\frac{1}{\sigma_H^2} - \frac{1}{\sigma_L^2}\right) > 0$, we have $\frac{\partial \hat{P}(\sigma_L^2|x)}{\partial \alpha} > 0$, so as hindsight bias increases ($\alpha$ increases), $\hat{P}(\sigma_L^2 \mid x)$ increases. Thus,

$$\hat{P}(\sigma_L^2 \mid x) > P(\sigma_L^2 \mid x) \text{ and } \hat{P}(\sigma_H^2 \mid x) < P(\sigma_H^2 \mid x). \tag{19.31}$$

In finance, volatility is often defined as the standard deviation of returns. From (19.31), decision makers who are subject to hindsight bias infer, ex-post, that volatility is likely to be low rather than high. Furthermore, the degree of underestimation of volatility is increasing in the degree of hindsight bias, as captured by the size of $\alpha$. Since hindsight-biased decision makers underestimate volatility, if they have to invest in the relevant asset/portfolio, they will invest more than they really should, particularly if they are risk averse. One may expect that the performance of hindsight-biased individuals will, therefore, be inferior to those who are not hindsight-biased.

Using a within-subjects design, Biais and Weber (2009) propose the following empirical measure of hindsight bias: $\phi = \frac{\hat{\mu}-\mu}{x-\mu}$. Using (19.25), we get that $\phi = \alpha$. They consider two treatments. In the first treatment, the experimenter informs subjects, ex-post, of their ex-ante prediction, hence, eliminating the hindsight bias. In the second treatment, the hindsight bias is not eliminated. Subjects were given the market price of five German stocks (BASF, IKB, EON, Postbank, Premiere), two commodities (oil and gold), and the euro–dollar exchange rate. They were then asked to predict the prices for one week ahead. Subjects were also asked to give the upper and lower bounds of prices, which were converted into standard deviations or *initial implied volatility estimates*. A week later, once the relevant prices had been realized, the first treatment group was simply reminded of their previous estimates of prices and volatility, while the second treat-

ment group was asked to make the postdictive judgments (i.e., give a best estimate of their earlier estimates).

The results are shown in Table 19.7. The first row shows the hindsight bias calculated in the second treatment (this is $\phi = \alpha$). There is a positive hindsight bias for all eight assets, which is significant at the 10% level for four of the assets. The second row calculates the ratio of their initial implied volatility estimates and the remembered volatility. For five of the eight assets, the median estimates exceed 1 and for four assets this is significant at the 10% level. For one asset, the ratio is slightly less than 1 but it is statistically insignificant.

Hindsight-biased decision makers appear not to be surprised because they act as if they knew it all along. To test this, the *true surprise*, based on the ex-ante measures is defined as $\left(\frac{x-\mu}{\sigma}\right)^2$, where $\sigma$ is the initial implied volatility. Define *biased surprise* as $\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right)^2$, where $\hat{\mu}$ and $\hat{\sigma}$ are remembered mean and remembered volatility; this measure can only be computed in treatment 2. For treatment 2, the biased surprise is smaller than the true surprise; the difference is statistically significant for two cases—Gold and dollar.

At the second date, the subjects are asked, yet again, to forecast their volatility estimates for another week ahead. The upward revisions in volatility are stronger for treatment 1 subjects relative to treatment 2 subjects. This is consistent with hindsight-biased subjects not updating volatility estimates as much as they ideally should, in the absence of the bias.

In order to test the performance of hindsight-biased actual investors, data was used from the Frankfurt and London branches of a large investment bank. A between-subjects design was used because competitive traders might not wish to admit that they forecast the future in a biased manner. The questions asked were similar in nature to those that the traders encountered in their jobs. Traders in both offices were found to be almost equally hindsight-biased. The estimate of $\alpha$ in the London and Frankfurt offices was, respectively, 0.547 and 0.579. Traders are split into high, middle, and low earners. Those in the high earners category demonstrate a lower mean and median hindsight bias, relative to the other two categories. The median difference between the middle- and low-income categories is similar. For individual-level data, and controlling for age and experience, the very highest earners have the lowest hindsight bias.

## 19.9 Confirmation bias

Suppose that an individual holds some initial beliefs. If the individual interprets subsequent evidence in a manner that is biased towards supporting the initial beliefs, relative to a Bayesian, then the individual is said to exhibit *confirmation bias*. For instance, buyers of a new car find that the particular model they drive suddenly appears more numerous on the roads. Since this takes the form of a biased assimilation of information, another name for this phenomenon, often used in psychology, is *biased assimilation*; for a survey of psychological research on the phenomenon see Nickerson (1998), Lord and Taylor (2009), and Mercier and Sperber (2011).

Rabin and Schrag (1999) note that confirmation bias is more likely to arise when: (1) the evidence is ambiguous and bears alternative interpretations, (2) the collection and scrutinization of evidence is selective (*hypothesis-based filtering*), (3) people are asked to judge the correlation between temporally separated events, and (4) individuals need to aggregate information from various sources. Confirmation bias is not mitigated by greater cognitive abilities or open mindedness (Stanovich and West 2007, 2008a, b).

A natural implication of confirmation bias is that individuals may be too overconfident because they typically search for confirmatory evidence and ignore or underweight contradictory evidence.

**Table 19.7** Median values across decision makers. p-values are given in parentheses.

| | BASF | IKB | EON | Postbank | Premiere | Oil | Gold | Dollar |
|---|---|---|---|---|---|---|---|---|
| Hindsight bias second treatment | 0.086 | 0.463 | 0.155 | 0.784 | 0.375 | 0.182 | 0.094 | 0.264 |
| | (0.344) | (0.009) | (0.166) | (0.000) | (0.626) | (0.074) | (0.052) | (0.108) |
| Initial implied $\sigma$/remembered | 1.142 | 1.267 | 1.000 | 1.577 | 1.273 | 1.090 | 1.000 | 0.944 |
| Implied $\sigma$ second treatment | (0.082) | (0.002) | (0.822) | (0.000) | (0.000) | (0.087) | (0.517) | (0.525) |
| True surprise first treatment | 2340 | 25 | 57004 | 737 | 12 | 3320 | 418167 | 1.5 |
| True surprise second treatment | 3026 | 22 | 34649 | 717 | 18 | 5867 | 226082 | 1.1 |
| Rank sum test difference between both treatments | (0.969) | (0.787) | (0.150) | (0.435) | (0.817) | (0.777) | (0.164) | (0.472) |
| Biased surprise second treatment | 2073 | 8 | 21007 | 215 | 38 | 5.504 | 187809 | 0.7 |
| Rank sum test difference between true and biased surprise | (0.910) | (0.667) | (0.158) | (0.866) | (0.278) | (0.765) | (0.024) | (0.041) |
| Initial implied $\sigma$/subsequent implied $\sigma$ first treatment | 0.923 | 1.024 | 0.630 | 1.001 | 1.041 | 0.981 | 0.996 | 0.892 |
| Initial implied $\sigma$/subsequent implied $\sigma$ second treatment | 0.978 | 1.373 | 0.858 | 1.854 | 1.407 | 1.003 | 0.846 | 0.910 |
| Rank sum test for difference between the two treatments | (0.172) | (0.021) | (0.099) | (0.002) | (0.001) | (0.349) | (0.520) | (0.369) |

Source: Biais and Weber (2009).

Confirmation bias may also lead to the formation and hardening of prejudices and stereotypes. In the presence of confirmation bias, learning may lead asymptotically to the incorrect model of the world with probability 1. Rabin and Schrag (1999) distinguish confirmatory bias from a related phenomenon in which once subjects form strong beliefs about one hypothesis over competing hypotheses, they start being *inattentive* to new evidence for or against their preferred hypothesis.

The nature of confirmatory bias is beautifully captured in the following quote from Lord et al. (1979, p. 2099):

> Thus, there is considerable evidence that people tend to interpret subsequent evidence so as to maintain their initial beliefs. The biased assimilation processes underlying this effect may include a propensity to remember the strengths of confirming evidence but the weaknesses of disconfirming evidence, to judge confirming evidence as relevant and reliable but disconfirming evidence as irrelevant and unreliable, and to accept confirming evidence at face value while scrutinizing disconfirming evidence hypercritically. With confirming evidence, we suspect that both lay and professional scientists rapidly reduce the complexity of the information and remember only a few well-chosen supportive impressions. With disconfirming evidence, they continue to reflect upon any information that suggests less damaging "alternative interpretations." Indeed, they may even come to regard the ambiguities and conceptual flaws in the data opposing their hypotheses as somehow suggestive of the fundamental correctness of those hypotheses. Thus, completely inconsistent or even random data—when "processed" in a suitably biased fashion—can maintain or even reinforce one's preconceptions.

### 19.9.1 *Empirical evidence for confirmation bias*

The typical experiments in confirmatory bias proceed as follows. The initial preference of subjects, or prior beliefs, for one hypothesis over another is elicited. For instance, do you support capital punishment or not (Lord et al., 1979); do you think nuclear technology is safe or unsafe (Plous, 1991); which football team do you support (Hastorf and Cantril, 1954); which party do you support in the US Presidential elections (Munro et al., 2002). The subjects are then given some information or evidence that supports one hypothesis over another and asked to form posterior beliefs. Confirmation bias arises when, relative to a Bayesian, the posterior beliefs of subjects adjust insufficiently from their priors. Such a bias was found in all the studies mentioned in this paragraph.

The initial beliefs of subjects in experiments can be created in the lab by providing them with anchoring information. Darley and Gross (1983) create two different profiles of a girl in two different treatments in a between-subjects design. In the first treatment, she is the daughter of college graduates who hold white collar jobs; a video clip shows the girl playing in a prosperous, middle class neighborhood. In a second treatment, she is the daughter of high school graduates who hold blue collar jobs; a video clip shows the girl playing in a poor, inner-city neighborhood. The subjects are divided into two sets; set A is used to elicit prior beliefs and set B to elicit posterior beliefs. Based on the information provided, the average estimate of set A individuals about the reading abilities of the girl (prior beliefs), in the first treatment, is slightly higher (4.29) as compared to the second treatment (3.90).

Subjects in set B are provided some evidence of the girl's scholastic abilities; this takes the form of a video clip in which the girl answers a series of questions with varying degrees of success. The average posterior beliefs of individuals in set B about the girl's reading abilities, in both treatments, were significantly different from the prior beliefs of set A. For instance, set B subjects in the second treatment (inner-city) rated the girl's average ability to be 3.71 (posterior belief), which

is below the prior estimate of set A individuals, 3.90. The corresponding average posterior estimate of set B individuals in the first treatment (middle class) was 4.67, which is statistically higher than the prior estimate of set A individuals, 4.29. Despite the fact that subjects viewed identical videos, the girl's background has a significant influence on her perceived scholastic abilities. Girls from poor inner-city neighborhoods are assigned even lower ability levels while girls from more prosperous neighborhoods are assigned even higher ability levels. Thus, subjects exhibit the confirmation bias.

## 19.9.2 *A formal model of confirmation bias and overconfidence*

Rabin and Schrag (1999) propose a stylized model of confirmation bias that nicely illustrates the concept of confirmation bias and shows how it might lead to overconfidence. We consider this model below. Suppose that there are two mutually exclusive and exhaustive states of the world, $x \in \{A, B\}$; $A$ and $B$ may be thought of as competing hypotheses concerning a particular issue/event that an individual wishes to form beliefs about. The prior probability is $p(A) = p(B) = 0.5$. Time is discrete and denoted by $t \in \{1, 2, 3, \ldots\}$.

At time $t$, the individual receives a signal, $s_t \in \{a, b\}$, that is correlated with the true state, $A$ or $B$, of the world. Let $p(a \mid A)$ be the probability of receiving signal $a$ if hypothesis $A$ is true. Similarly, $p(b \mid B)$ is the probability of signal $b$ if hypothesis $B$ is true. The signals are temporally *iid* and $p(a \mid A) = p(b \mid B) = \theta$ and $0.5 < \theta < 1$. If the underlying hypothesis $A$ is true, then $p(a \mid A) = \theta$ and $p(b \mid A) = 1 - \theta$; likewise if $B$ is true then $p(b \mid B) = \theta$ and $p(a \mid B) = 1 - \theta$. After observing the signal, the individual uses Bayes' rule to update the beliefs about the relative likelihood of $x = A$ and $x = B$.

Confirmatory bias takes the form of misinterpreting signals that are in conflict with one's current beliefs about which of the two hypotheses is likely to be true. The signals perceived by the individual are given by $\sigma_t \in \{\alpha, \beta\}$. When $\alpha$ (respectively, $\beta$) is the perceived signal, the individual believes that the received signal is $a$ (respectively, $b$). Signals that confirm the initial beliefs are correctly encoded, while signals that conflict with the initial beliefs are incorrectly perceived with probability $q > 0$.

Suppose, for instance, that at time $t$, the individual believes $A$ is more likely than $B$. If the signal is $s_t = a$, then this is correctly encoded, i.e., $\sigma_t = \alpha$. But if $s_t = b$, then this is incorrectly perceived as $\sigma_t = \alpha$ with probability $q$ and $\sigma_t = \beta$ with probability $1 - q$. Similarly if $B$ is initially perceived to be more likely, then the signal $s_t = b$ is correctly perceived ($\sigma_t = \beta$) but $s_t = a$ is incorrectly perceived as $\sigma_t = \beta$ with probability $q$.

Confirmatory bias implies that although $s_t$ is temporally *iid*, $\sigma_t$ is not temporally *iid*. Denote the history of signals at the beginning of time $t$ by $s^t = \{s_1, \ldots, s_{t-1}\}$ and the history of perceived signals at the beginning of $t$ by $\sigma^t = \{\sigma_1, \ldots, \sigma_{t-1}\}$. Let $p(A \mid \sigma^t)$ and $p(B \mid \sigma^t)$ denote, respectively, the beliefs of the individual at the beginning of time $t$ about hypotheses $A$ and $B$.

Let $\theta^F$ be the probability that the individual perceives a signal $\sigma_t$ that confirms his hypothesis, when it is *false* (i.e., the opposite hypothesis is, in fact, true), then

$$\theta^F = \begin{cases} p(\sigma_t = \alpha \mid p(A \mid \sigma^t) > 0.5, B) \\ p(\sigma_t = \beta \mid p(B \mid \sigma^t) > 0.5, A) \end{cases}, \tag{19.32}$$

where the conditioning expression $p(A \mid \sigma^t) > 0.5, B$ means that, conditional on the history of perceived signals, hypothesis $A$ is thought to be more likely, when, in fact, $B$ is true. Similarly the conditioning expression $p(B \mid \sigma^t) > 0.5, A$ means that conditional on the history of perceived signals, hypothesis $B$ is thought to be more likely, when, in fact, $A$ is true.

Consider the first row in (19.32). Suppose that conditional on the history of perceived signals at time $t$, $\sigma^t$, the individual believes $A$ to be more likely, i.e., $p(A \mid \sigma^t) > 0.5$. How can he perceive $\sigma_t = \alpha$ when $B$ is actually the correct hypothesis? Since $B$ is the correct hypothesis, $p(b \mid B) = \theta$ and $p(a \mid B) = 1 - \theta$. If $s_t = a$, which, in this case, occurs with probability $1 - \theta$, the individual's beliefs are confirmed, so he perceives it correctly. However, with probability $\theta$, the signal is $b$, which contradicts the initial beliefs about model $A$, so it is incorrectly perceived as $\alpha$ with probability $q$, hence

$$\theta^F = (1 - \theta) + \theta q. \tag{19.33}$$

Since the model is symmetric, the second row of (19.32) also leads to the same calculation reported in (19.33).

Let $\theta^T$ be the probability that the individual perceives a signal $\sigma_t$ that confirms his hypothesis, when it is *true* (i.e., the opposite hypothesis is false), then

$$\theta^T = \begin{cases} p(\sigma_t = \alpha \mid p(A \mid \sigma^t) > 0.5, A) \\ p(\sigma_t = \beta \mid p(B \mid \sigma^t) > 0.5, B) \end{cases}. \tag{19.34}$$

Consider the first row in (19.34). Since $A$ is the correct model in this case, $p(a \mid A) = \theta$ and $p(b \mid A) = 1 - \theta$; furthermore, the individual believes in model $A$ over $B$, $p(A \mid \sigma^t) > 0.5$. The perception that $\sigma_t = \alpha$ can come about in two ways. First, if the signal is $a$ (a probability $\theta$ event) it is perceived without error. Second, if the signal is $b$ (a probability $1 - \theta$ event), it is perceived mistakenly as $\alpha$ with probability $q$. Putting these together we get

$$\theta^T = \theta + (1 - \theta) q. \tag{19.35}$$

Symmetry ensures that the second row in (19.34) leads to an identical calculation.

**Definition 19.1** *(Unbiased Bayesian statistician): An unbiased Bayesian statistician has $q = 0$, so $\theta^F = (1 - \theta)$ and $\theta^T = \theta$.*

**Definition 19.2** *(Confirmation-biased individual): An extreme confirmation-biased individual has $q = 1$, so $\theta^F = 1$ and $\theta^T = 1$. Such an individual always misreads any contradictory evidence, hence, takes all evidence as supportive of his initial beliefs. Partial confirmation bias arises when $0 < q < 1$. In general, the degree of confirmation bias is increasing in $q$. All individuals who have $q \neq 0$ are said to be confirmation-biased.*

Let us superscript variables belonging to a confirmation-biased individual by $C$ and a Bayesian individual by $B$. Suppose that the individual observes $n$ signals; $n_\alpha$ are perceived to be $\alpha$ and $n_\beta$ are perceived to be $\beta$ ($n_\alpha + n_\beta = n$). What are the posterior beliefs of a confirmation-biased individual that the true model is $A$? Using Bayes' rule,

$$p\left(A \mid n_\alpha, n_\beta\right) = \frac{p\left(n_\alpha, n_\beta \mid A\right) p(A)}{p\left(n_\alpha, n_\beta \mid A\right) p(A) + p\left(n_\alpha, n_\beta \mid B\right) p(B)}. \tag{19.36}$$

The prior probabilities are $p(A) = p(B) = 0.5$. When the true model is $A$, the probability of $n_\alpha$ outcomes that are $\alpha$ and $n_\beta$ outcomes that are $\beta$ is $p\left(n_\alpha, n_\beta \mid A\right) = \theta^{n_\alpha} (1 - \theta)^{n_\beta}$. Similarly, we can compute $p\left(n_\alpha, n_\beta \mid B\right) = \theta^{n_\beta} (1 - \theta)^{n_\alpha}$. Substituting these calculations in (19.36), we get

$$p\left(A \mid n_\alpha, n_\beta\right) = \frac{\theta^{n_\alpha}(1-\theta)^{n_\beta}}{\theta^{n_\alpha}(1-\theta)^{n_\beta} + \theta^{n_\beta}(1-\theta)^{n_\alpha}}.$$

Similarly computing $p\left(B \mid n_\alpha, n_\beta\right)$, we can determine the likelihood ratio of a confirmation-biased individual,

$$L^C(n_\alpha, n_\beta) = \frac{p\left(A \mid n_\alpha, n_\beta\right)}{p\left(B \mid n_\alpha, n_\beta\right)} = \left(\frac{\theta}{1-\theta}\right)^{n_\alpha - n_\beta}. \tag{19.37}$$

Thus, conditional on the history, $n_\alpha, n_\beta$, the confirmation-biased individual believes that $A$ is more likely if $L^C(n_\alpha, n_\beta) > 1$ and $B$ is more likely if $L^C(n_\alpha, n_\beta) < 1$. Since we have assumed that $0.5 < \theta < 1$, whenever $n_\alpha > n_\beta$, a confirmation-biased individual will assign greater probability to $A$ being true relative to $B$. For instance, suppose that $n = 5$ and $n_\alpha = 3$, $n_\beta = 2$. Using (19.37), we have $L^C(n_\alpha, n_\beta) = \frac{\theta}{1-\theta} > 1$. Hence, a confirmation-biased individual believes relatively more in $A$.

Relative to a confirmation-biased individual, denote the likelihood ratio formed by an unbiased Bayesian statistician by $L^B(n_\alpha, n_\beta)$. Suppose that the hypothetical unbiased Bayesian statistician knows that an individual suffers from confirmation bias and also misperceives signals in the manner outlined above. In general, we will have $L^B(n_\alpha, n_\beta) \neq L^C(n_\alpha, n_\beta)$. Since the perceived signals $\sigma_t$ are not *iid* (although the actual signals $s_t$ are iid), the order of the perceived signals received by a confirmation-biased individual influence the posterior beliefs.

**Definition 19.3** *Suppose that $n_\alpha > n_\beta$. The confirmation-biased individual is:*

*(i) Overconfident if $L^B(n_\alpha, n_\beta) < L^C(n_\alpha, n_\beta)$. Such an individual believes that hypothesis A is more likely than can be justified by the evidence.*
*(ii) Underconfident if $L^B(n_\alpha, n_\beta) > L^C(n_\alpha, n_\beta)$.*

**Example 19.1** *(Overconfidence): Suppose that the sequence of perceived signals of the confirmation-biased individual is $\alpha, \alpha, \alpha, \beta, \beta$. What likelihood should an unbiased Bayesian statistician assign to A versus B? We assume that our hypothetical unbiased Bayesian statistician knows that the individual suffers from confirmation bias and also misperceives signals in the manner outlined above.*
*The Bayesian likelihood ratio is*

$$L^B = \frac{p\left(A \mid \alpha, \alpha, \alpha, \beta, \beta\right)}{p\left(B \mid \alpha, \alpha, \alpha, \beta, \beta\right)}.$$

*Using Bayes' rule*

$$p\left(A \mid \alpha, \alpha, \alpha, \beta, \beta\right) = \frac{p\left(\alpha, \alpha, \alpha, \beta, \beta \mid A\right) p(A)}{p\left(\alpha, \alpha, \alpha, \beta, \beta \mid A\right) p(A) + p\left(\alpha, \alpha, \alpha, \beta, \beta \mid B\right) p(B)}.$$

*Writing the analogous expression for $p\left(B \mid \alpha, \alpha, \alpha, \beta, \beta\right)$ and noting that $p(A) = p(B) = 0.5$ we have*

$$L^B = \frac{p\left(\alpha, \alpha, \alpha, \beta, \beta \mid A\right)}{p\left(\alpha, \alpha, \alpha, \beta, \beta \mid B\right)}. \tag{19.38}$$

*Let us first compute $p\left(\alpha, \alpha, \alpha, \beta, \beta \mid A\right)$, i.e., what is the likelihood of getting the five perceived signals in the order given, conditional on the true model being A. Given the signals, since*

*$n_\alpha > n_\beta$, the individual always believes the true model is more likely to be A. Conditional on the true model being A (or $x = A$), the probability of the first $\alpha$ signal is $\theta$. The second period signal could have been either $s_t = a$ (with probability $\theta$) and correctly perceived to be $\sigma_t = \alpha$ or it might have been $s_t = b$ (with probability $1-\theta$) and incorrectly perceived to be $\sigma_t = \alpha$ with probability $q$. This calculation is $\theta^T = p(\alpha \mid p(A \mid \sigma^t) > 0.5, A)$, the first line in (19.34). The third period perceived signal is $\alpha$ and gives rise to an identical calculation as in period 2. In period 4, the perceived signal is $\beta$. Now the relevant calculation is $1-\theta^T$. An identical calculation is repeated in period 5. Thus,*

$$\begin{aligned} p(\alpha,\alpha,\alpha,\beta,\beta \mid A) &= \theta\left(\theta^T\right)^2\left(1-\theta^T\right)^2 \\ &= \theta\left(\theta+(1-\theta)q\right)^2(1-\theta)^2\left(1-q\right)^2 \ \textit{(using (19.35)).} \end{aligned} \tag{19.39}$$

*Proceeding analogously, $p(\alpha,\alpha,\alpha,\beta,\beta \mid B) = (1-\theta)\left(\theta^F\right)^2\left(1-\theta^F\right)^2$. Substituting the value of $\theta^F$ from (19.33) we get*

$$p(\alpha,\alpha,\alpha,\beta,\beta \mid B) = (1-\theta)\left((1-\theta)+\theta q\right)^2\theta^2\left(1-q\right)^2. \tag{19.40}$$

*Substituting (19.39), (19.40) in (19.38), the Bayesian likelihood ratio is*

$$\begin{aligned} L^B &= \frac{\theta\left(\theta+(1-\theta)q\right)^2(1-\theta)^2\left(1-q\right)^2}{(1-\theta)\left((1-\theta)+\theta q\right)^2\theta^2\left(1-q\right)^2}, \\ &= \left(\frac{\theta+q(1-\theta)}{1-\theta+q\theta}\right)^2\frac{1-\theta}{\theta}. \end{aligned}$$

*It is now straightforward to show that $\left(\frac{\theta+q(1-\theta)}{1-\theta+q\theta}\right)^2\frac{1-\theta}{\theta} < \frac{\theta}{1-\theta}$, hence, $L^B < L^C$. Thus, the confirmation-biased individual is overconfident.*

**Example 19.2** *Suppose that the sequence of perceived signals of the confirmation-biased individual is $\beta,\beta,\alpha,\alpha,\alpha$. In this case, it can be shown that*

$$L^B = \frac{(1-\theta)\left(1-\theta+q\theta\right)\theta^3}{\theta\left(\theta+q(1-\theta)\right)(1-\theta)^3} > \frac{\theta}{1-\theta} = L^C. \tag{19.41}$$

*Hence, the individual is underconfident.*

## 19.10 Other judgment heuristics

We now briefly note some other judgment heuristics in this section. The aim is to provide no more than a minimum discussion of each heuristic without surveying the literature in any detail.

### 19.10.1 *Regression towards the mean*

Galton (1886) discovered that taller than average parents produce children who, on average, are shorter than them. Similarly, shorter than average parents produce children whose average height is greater than theirs. This phenomenon is known as *regression towards the mean*. Regression towards the mean always holds between two random variables whose correlation is different from 1 in absolute value.

Suppose that we measure test scores, or some other human performance measures, at two distinct points in time; each score is partly determined by random influences. Suppose that the test scores are random variables, $X_1$ and $X_2$, which are identically distributed with common mean $\mu$. Let us select from the first test scores, individuals whose scores deviate from the mean by more than $k > 0$, i.e., scores from the set $\{x : x - \mu > k, x \in X_1\}$. Suppose that we now track the performance of the same individuals on the second test where the test score is determined by the random variable $X_2$. Denote the sample mean of the test scores of these selected individuals in the second test by $\mu_s$. Then regression to the mean implies that $\mu_s - \mu < k$. Similarly, if we chose the original set of individuals from distribution $X_1$ who satisfy $\{x : \mu - x > k, x \in X_1\}$, then we must have $\mu - \mu_s < k$.

Consider the following interesting example of this phenomenon from the Israeli Air Force (Tversky and Kahneman, 1974; Kahneman, 2011). Air instructors found that when flight cadets were praised for particularly good performance, say a clean execution of an aerobatic maneuver, then their subsequent performance dropped. On the other hand, those flight cadets who were punished for poor performance, improved their subsequent performance. An inference was made that rewarding good performance was bad and punishing bad performance was good. However, this fits perfectly with an inability to appreciate the statistical phenomenon of regression to the mean.

On average, extreme performances on the aerobatic maneuver arise purely from luck. Flight cadets who were extremely lucky performed much better than the average while others did worse than the average. There is no causality between rewards/punishments and subsequent performance. The fluctuations in performance between two tests is purely random. It is of course true that low skilled pilots will continue to perform at a lower level, however, regression to the mean captures the variability in performance that arises from purely random factors.

The regression to the mean has been documented in other contexts. For instance, golfers with high scores on day 1 of a tournament typically report lower scores on day 2, while those who struggle on day 1, improve their average performances on day 2. Kahneman (2011) conjectures that people give inadequate attention to regression to the mean because the brain wishes to construct causal relations while regression is a purely statistical phenomenon that has an explanation but not a cause.

### 19.10.2 *False consensus effect*

The *false consensus* heuristic arises when people believe that their own behavioral choices and judgments are appropriate, and relatively common among others. Individuals walking on a tightrope across skyscrapers or launching a revolution do know that they are in a minority. Yet they also view their behavior as less deviant, relative to the behavior of others who do not engage in these behaviors (Ross et al., 1977).

The false consensus effect has some similarity with *evidential reasoning* or *projection bias*, and this is recognized in Ross et al. (1977). Under evidential reasoning, subjects ascribe diagnostic significance to their own beliefs when asked to guess the beliefs of others; see Part 4 of the book for a formalization of evidential reasoning (al-Nowaihi and Dhami, 2015), and the experimental evidence (Robbins and Krueger, 2005).

In one experiment, Ross et al. (1977) asked Stanford University students whether they would be willing to walk around the campus for 30 minutes wearing a sandwich board inscribed with a message. For one group of students, the message read "Eat at Joe's" and for another group it read "Repent." When the message was Repent, those who agreed, estimated, on average, that 63.50%

of their fellow students would also agree, and those who refused, estimated that 76.67% of their fellow students would also refuse. For both messages combined, those who agreed to wear the board estimated that 62.2% of other students would also agree to wear it. Those who refused to wear the message board estimated that 67% would also refuse.

For a meta-study of 115 studies on the false consensus effect, see Mullen et al. (1985). Solan et al. (2008) show that laymen and judges alike are subject to the false consensus effect when interpreting contractual documents in insurance that have in the past led to different judicial outcomes. Leviston et al. (2013) find that individuals believe that their own views on climate are more widespread than they actually are. Those who thought that climate change was not happening believed that 43% of others thought likewise; the actual figure was 6%. Similarly, those who said "don't know" thought that 34.1% of others believed likewise; the actual figure was less than 5%. Roth and Voskort (2014) document a false consensus effect for finance professionals in Germany. In an artefactual field experiment, finance professionals are asked to guess the risk preferences of others. A significant false consensus effect is found and professionals unduly project their own risk preferences to others. Indeed, the measured effect is strongest for experienced professionals.

### 19.10.3 *Confusion between necessary and sufficient conditions*

The distinction between *necessary and sufficient conditions* is critical for most inference in economics, yet people can find it difficult to make the distinction. Consider the following classic problem in Wason (1968). There are four cards labeled E, K, 4, 7 that are placed on a table. It is known that each card has a number on one side and a letter on the other. Subjects are invited to pick all the cards that can test the following rule. "*Every card with a vowel on one side has an even number on the other side*."

Most subjects report that card E must be turned over while many report that both E and 4 must be turned over to test the rule. The rule can be written as the following logical statement:

$$\text{Vowel on one side} \Rightarrow \text{even number on the other side.} \tag{19.42}$$

Thus, a vowel on one side is sufficient for there to be an even number on the other side. But an even number on one side is only a necessary, but not a sufficient condition for a vowel on the other side. An implication of (19.42) is that

$$\text{not an even number on one side} \Rightarrow \text{not a vowel on the other side.} \tag{19.43}$$

Statement (19.42) dictates choosing card E, and (19.43) dictates choosing card 7. Choosing card 4 can never confirm (or falsify) the rule because an even number is not a sufficient condition for a vowel on the other side, yet about 65% choose card 4.

### 19.10.4 *Attribute substitution*

Kahneman and Frederick (2005) draw attention to the heuristic of *attribute substitution*. This occurs when one attribute of an object is substituted by another, perhaps because it comes more readily to mind. For instance, "similarity" might be substituted for "probability" in solving a prob-

lem. The usefulness of attribute substitution is that it simplifies difficult judgments by substituting easier ones in their place. The downside is, of course, that it may lead to incorrect inferences.

Consider, for instance, the following two questions asked of subjects in experiments (Strack et al., 1988). How happy are you with your life in general? How many dates did you have last month? When asked in this order, the correlation in the answers was low. However, when the order was reversed, the correlation rose to 0.66. Kahneman and Frederick (2005) propose the following explanation: "We suggest that thinking about the dating question automatically evokes an affectively charged evaluation of one's satisfaction in that domain of life, which lingers to become the heuristic attribute when the happiness question is subsequently encountered." Schwarz, Strack, and Mai (1991) replicate these results when marital satisfaction replaces the number of dates (the respective correlations were 0.32 when the questions were asked in direct order, and 0.67 in the reversed order). For a review of findings from similar experiments, see Schwarz and Bohner (2001).

## 19.11 Dual process models and judgment heuristics

In this section, we briefly outline some of the explanations for judgment heuristics and biases. Although our treatment is relatively condensed, this is an important area of research in psychology that has attracted a great deal of scholarship.[17] Various judgment heuristics can be explained by taking a *two modules view* of the brain (Kahneman, 2011). The two modules have several distinct names in the literature such as "dual process theories" and "intuition and reason." Following Stanovich and West (2000, 2002), we refer to the two modules as System 1 and System 2; these might not be the most illuminating terms but they are easiest to use and are popular.

System 1 is fast, automatic, and little mental effort is involved in using it. System 2, on the other hand, is much slower, deliberative, and requires cognitive effort. System 1 forms impressions, feelings, and intuitions for System 2. If System 2 endorses these intuitions, they turn into beliefs. System 2 typically springs into action when System 1 encounters a situation that is counter to its worldview. Table 19.8 gives the different usages of these terms, which illustrate nicely the distinction between the two systems, and the various features associated with them.

Kahneman (2011, p. 25) offers a very illuminating view of the two systems:

> The division of labor between System 1 and System 2 is highly efficient: it minimizes effort and optimizes performance . . . System 1 is generally very good at what it does: its models of familiar situations are accurate, its short term predictions are accurate as well, and its initial reactions to challenges are swift and generally appropriate. System 1 has biases, however, and it is prone to making systematic errors in specified circumstances . . . it answers easier questions than the one it was asked and has little understanding of logic and statistics. One further limitation of System 1 is that it cannot be turned off.

It is important to note that the two systems are purely fictitious characters that provide a way of organizing our understanding of the brain. Indeed "there is no part of the brain that either of the systems would call home" (Kahneman, 2011, p. 29).

One explanation for judgment heuristics is that individuals save on cognitive effort by using the relatively less cognitively taxing and spontaneous System 1, as compared to System 2. This

[17] The distinction between Systems 1 and 2 that we outline in this section is used extensively in formal economic models; see Part 6 of the book.

**Table 19.8** Different usages of the System 1–System 2 terms.

| | System 1 | System 2 |
|---|---|---|
| Dual-Process Theories: | | |
| Sloman (1996) | associative system | rule-based system |
| Evans (1984; 1989) | heuristic processing | analytic processing |
| Evans & Over (1996) | tacit thought processes | explicit thought processes |
| Reber (1993) | implicit cognition | explicit learning |
| Levinson (1995) | interactional intelligence | analytic intelligence |
| Epstein (1994) | experiential system | rational system |
| Pollock (1991) | quick and inflexible modules | intellection |
| Hammond (1996) | intuitive cognition | analytical cognition |
| Klein (1998) | recognition-primed decisions | rational choice strategy |
| Johnson-Laird (1983) | implicit inferences | explicit inferences |
| Shiffrin & Schneider (1977) | automatic processing | controlled processing |
| Posner & Snyder (1975) | automatic activation | conscious processing system |
| Properties: | associative | rule-based |
| | holistic | analytic |
| | automatic | controlled |
| | relatively undemanding of cognitive capacity | demanding of cognitive capacity |
| | relatively fast | relatively slow |
| | acquisition by biology, exposure, and personal experience | acquisition by cultural and formal tuition |
| Task Construal | highly contextualized | decontextualized |
| | personalized | depersonalized |
| | conversational and socialized | asocial |
| Type of Intelligence Indexed: | interactional (conversation and implicature) | analytic (psychometric IQ) |

Source: Keith E. Stanovich and Richard F. West (2000). "Individual differences in reasoning: implications for the rationality debate?" *Behavioral and Brain Sciences* 23: 645–665, with permission by Cambridge University Press.

is also known as the *cognitive miser* model. For instance, Tversky and Kahneman (1973) found that mathematical psychologists exhibited the law of small numbers. Despite being well trained in statistics, they used their fast and reactive System 1, thus neglecting basic statistical analysis.

System 1 is biased to look for confirmatory evidence that enables individuals to make sense of the world around them. Such a bias is also reflected in the desire of individuals to reduce or avoid *cognitive dissonance*. Cognitive dissonance arises from the psychological stress placed upon individuals when they hold different *mental models* of the world (see Section 19.13 on mental models). The individuals' response in such cases is to rationalize the world around them and reduce conflict between different mental models. The term "cognitive dissonance" was proposed by Festinger (1957) and a classic set of experiments can be found in Festinger and Carlsmith (1959).

The fact that the suppression of an incorrect response by System 1 requires cognitive effort on the part of System 2 can be illustrated by the *Stroop task*. The term Stroop task is now used for any generic task in which an automatic, highly practiced response, by System 1 is, in fact, incorrect. System 2 then attempts to control the incorrect response (successfully or unsuccessfully), which accounts for the hesitation in subject responses. In the classical example, subjects are asked to read names of different colors, but some of the names are printed in a different color, e.g., the word "green" may be printed in the color "red." Subjects take longer to answer these questions and make more errors relative to the case where the names of the colors and the color in which they are printed, match.

System 1 is also associated with high levels of skill and achievement. For instance, expert chess players can quickly gauge the strength of their position without considering all possibilities. Indeed, considering all possibilities would be seriously taxing even for System 2 and deplete scarce cognitive resources. Kahneman and Frederick (2005) provide a list of factors that determine the relative usage of the two systems that we now outline. Unfamiliar tasks and those involving abstract concepts are likely to be dealt with by System 2. However, System 1 is likely to be used to a greater extent when the time available for deliberation is smaller, exposure to statistical training is lower, and some traditional measures of IQ are lower.

The relative use of the two systems might also depend on the moods of individuals. A representation of the data in terms of frequencies, rather than probabilities, can also influence the ability of System 2 to correct the errors of System 1. For instance, Kahneman and Frederick (2005) write: "the language of frequencies improves respondent's ability to impose the logic of set inclusion on their considered judgments." Indeed, we have reviewed a great deal of evidence above that in some cases, frequency information rather than probability information reduces some behavioral biases but the biases do remain, often for a majority of the subjects.

Higher cognitive ability might be associated with a greater and more efficient use of System 2. If this insight is correct, then we should see a lower prevalence of judgment errors among subjects in experiments whose cognitive ability is higher. Oechssler et al. (2009) test this idea. In order to measure cognitive ability, they use the *cognitive reflection test* (CRT) proposed in Frederick (2005). In this test, subjects are asked the following three questions.

1. A bat and a ball together cost 110 cents. The bat costs 100 cents more than the ball. How much does the ball cost?
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

If one only engages System 1, then the answers to the three questions are respectively 10 cents, 100 min, 24 days. However, these answers are incorrect. The correct answers based on engaging System 2 are, respectively, 5 cents, 5 min, 47 days. The average number of correct responses was 2.05.[18] As in Frederick's (2005) sample, a lower CRT score is associated with greater risk seeking and greater impatience (higher discount rate).

There is a statistically significant difference in the error rates of individuals with high and low CRT scores. For the conjunction fallacy (the Linda problem), 62.6% of the low CRT group engaged in the fallacy, while only 38.3% of the high CRT group did. Now consider the conservatism problem, or the underweighting of the likelihood of a sample. In a problem given

[18] In Frederick's (2005) sample this score lies between the scores for the MIT and Princeton students.

to the subjects, the probability consistent with Bayes' rule was 0.967; however, the low CRT group reported a probability of 0.58, while the high CRT group reported a probability of 0.69. Interestingly cognitive ability did not turn out to have any effect on the anchoring heuristic. This suggests that the relative extent to which the heuristics are hardwired in the brain, differs. Some heuristics can be reduced with greater cognitive effort/ability, while others cannot be.

Bergman et al. (2010) find that as cognitive ability increases, the anchoring effect reduces but does not vanish. One reason for the difference in their findings from Oechssler et al. (2009) is that they do not use the CRT but rather a cognitive ability test (CAT) that is based on answering 44 questions in 20 minutes. Dohmen et al. (2009) find, using data from the German population, that cognitive ability and the number of years of schooling are positively related to better performance in probability judgment tasks, and a reduced incidence of the gambler's fallacy. For the relation between IQ and the extent of judgment biases, see Stanovich (1999) and Stanovich and West (2002).

A higher IQ can also be a doubled edged sword. On the one hand, a higher IQ can induce a more efficient system 2 that corrects for the mistakes of system 1. However, on the other hand, a high IQ can induce individuals to focus on a more plausible error rather than a random error; see for instance, Kahneman (2000).

## 19.12 Coarse thinking and persuasion

*Persuasion* is central to determining the outcomes in several forms of human interaction. Activities associated with persuasion include advertising, political campaigns, religious preaching, selling activities, media campaigns, and military propaganda. Of these, advertising as a form of persuasion, is the one that economists have focused on the most. A dominant strand in industrial organization insists that the objective of advertising is the dissemination of information (Stigler, 1987; Grossman and Hart, 1980; Milgrom, 1981).

In contrast, in psychology, persuasion is understood as the transference of information from one context to another (Kahneman and Tversky, 1982; Zaltman, 1997). Consider the following three examples of successful persuasion, taken from Shleifer et al. (2008).

1. American Express travelers cheques featured Karl Malden, an actor who played a detective in a well-known television show. The advertisements suggested that, for safety, people should carry American Express travelers cheques in preference to cash.
2. After the crash of the Internet bubble, the brokerage firm Charles Schwab launched an advertising campaign that featured judges, ship captains, crossing guards, doctors, grandmothers, and other steady and reliable individuals, each standing next to a Schwab investment specialist. The headline of one advertisement, depicting a Schwab professional standing next to a pediatrician, said: "Both are seen as pillars of trust."
3. Consider Arnold Schwarzenegger's memorable speech at the 2004 Republican National Convention. In the best remembered part of his speech, Schwarzenegger defended free trade: "To those critics who are so pessimistic about our economy, I say: Don't be economic girlie men! . . . Now they say India and China are overtaking us. Don't you believe it. We may hit a few bumps—but America always moves ahead. That's what Americans do."

All these examples illustrate a kind of thinking, known as *associative*, *analogical*, *or metaphoric thinking* in psychology. Information that has great relevance in one context is used in another context where it has much less, or worse, no relevance at all. Based on Mullainathan (2002),

Shleifer et al. (2008) make progress in formalizing these ideas by focusing on the concept of *coarse thinking*.

To explain *coarse thinking*, suppose that individuals are able to categorize a problem into one of $n$ mutually exclusive categories. Any two problems in the same category have identical methods of solution. So, for instance, if there are two distinct problems in category $j \in \{1, \ldots, n\}$, then they are solved using an identical model. The polar opposite of coarse thinking is known as *sophisticated thinking* in which different models may be suitable for each problem. Coarse thinking can be exploited by others for their benefit. So, for instance, a beautiful model or a well-known sports personality posing next to a car gives no information about the car, per se. However, individuals might, using associative thinking, transfer attributes they associate with the individual posing with the car, to the car itself.

Suppose that for some good, service, or information, a buyer needs to assess the level of quality, $q \in (-\infty, \infty)$. The buyer receives two possible messages $m \in \{\underline{m}, \overline{m}\}$, where $\underline{m}$ and $\overline{m}$ have the respective connotations of *bad* and *good* messages. The individual knows that there are two mutually exclusive *categories* or *situations* $s \in \{\underline{s}, \overline{s}\}$. Messages are only informative when $s = \overline{s}$; when $s = \underline{s}$ messages are completely uninformative. The joint distribution of quality, $q$, messages, $m$, and situations, $s$, is given by $p(q, m, s)$.

**Example 19.3** *Suppose that putting "silk" in shampoo has no effect on hair. Yet a shampoo company advertises putting silk in shampoo in the hope that associative thinking will persuade consumers to buy the shampoo. In this case, the message $\overline{m}$ = silky, while $\underline{m}$ is any neutral message. One situation $\overline{s}$ is that a person is looking for a shampoo, and in a second situation $\underline{s}$ the person is not looking for a shampoo but simply assessing the quality of the plastic used in shampoo bottles; so $\overline{s}$ = shampoo and $\underline{s}$ = bottle. The message $\overline{m}$ = silky is uninformative in the situation $\underline{s}$ but informative in the situation $\overline{s}$. A Bayesian thinker will simply avoid the shampoo ad in the first situation. However, a coarse thinker (see formal definition below) who lumps situations together in categories will give at least some weight to the shampoo advertisement.*

**Example 19.4** *Consider the first of three examples of successful persuasion given above on American Express travelers checks. Here the message $\overline{m}$ = Karl Malden advertisement, while $\underline{m}$ is any neutral message. The situations are: $\overline{s}$ = an individual with travel plans and $\underline{s}$ = an individual who has no travel plans. Clearly the message on travelers cheques is only relevant to individuals with travel plans.*

The assumptions are formally stated as follows.

**A1:** The distribution of quality, $q$, is independent of the situation, i.e.,

$$p(q \mid \underline{s}) = p(q \mid \overline{s}) = p(q). \tag{19.44}$$

**A2:** The average quality, $E\left[q\right] = 0$; using (19.44) this implies

$$E\left[q \mid s\right] = 0. \tag{19.45}$$

**A3:** Any message is possible under any situation

$$p(m \mid s) > 0;\ m \in \left\{\underline{m}, \overline{m}\right\},\ s \in \left\{\underline{s}, \overline{s}\right\}. \tag{19.46}$$

**A4:** Messages are informative only in situation $s = \overline{s}$. In situation $\overline{s}$, the message $\overline{m}$ is good news about average quality, while the message $\underline{m}$ is bad news about average quality. Thus,

$$E\left[q \mid \underline{m}, \overline{s}\right] < E\left[q \mid \overline{m}, \overline{s}\right]. \tag{19.47}$$

By contrast, in situation $\underline{s}$, the message is uninformative, i.e.,

$$p(q \mid \underline{m}, \underline{s}) = p(q \mid \overline{m}, \underline{s}) = p(q \mid \underline{s}) = p(q). \tag{19.48}$$

The last equality uses (19.44).

From the definition of conditional probability, we know that

$$p\left(q \mid \underline{m}, \overline{s}\right) = \frac{p\left(q, \underline{m}, \overline{s}\right)}{p\left(\underline{m}, \overline{s}\right)} = \frac{p\left(q, \underline{m}, \overline{s}\right)}{p\left(\underline{m} \mid \overline{s}\right) p\left(\overline{s}\right)}. \tag{19.49}$$

Analogous to (19.49), we can write

$$p\left(q \mid \overline{m}, \overline{s}\right) = \frac{p\left(q, \overline{m}, \overline{s}\right)}{p\left(\overline{m}, \overline{s}\right)} = \frac{p\left(q, \overline{m}, \overline{s}\right)}{p\left(\overline{m} \mid \overline{s}\right) p\left(\overline{s}\right)}. \tag{19.50}$$

From the definition of a marginal probability distribution, we get that

$$p\left(q, \underline{m}, \overline{s}\right) + p\left(q, \overline{m}, \overline{s}\right) = p\left(q, \overline{s}\right). \tag{19.51}$$

Substitute $p\left(q, \overline{m}, \overline{s}\right)$ from (19.50) in (19.51)

$$p\left(q, \underline{m}, \overline{s}\right) + p\left(q \mid \overline{m}, \overline{s}\right) p\left(\overline{m} \mid \overline{s}\right) p\left(\overline{s}\right) = p\left(q, \overline{s}\right). \tag{19.52}$$

Solve (19.52) for $p\left(q, \underline{m}, \overline{s}\right)$ and substitute in (19.49) to get

$$p\left(q \mid \underline{m}, \overline{s}\right) = \frac{p\left(q, \overline{s}\right)}{p\left(\underline{m} \mid \overline{s}\right) p\left(\overline{s}\right)} - \frac{p\left(q \mid \overline{m}, \overline{s}\right) p\left(\overline{m} \mid \overline{s}\right)}{p\left(\underline{m} \mid \overline{s}\right)}. \tag{19.53}$$

But $p\left(q \mid \overline{s}\right) = \frac{p\left(q, \overline{s}\right)}{p\left(\overline{s}\right)}$, hence, (19.53) can be written as

$$p\left(q \mid \underline{m}, \overline{s}\right) = \frac{p\left(q \mid \overline{s}\right)}{p\left(\underline{m} \mid \overline{s}\right)} - \frac{p\left(q \mid \overline{m}, \overline{s}\right) p\left(\overline{m} \mid \overline{s}\right)}{p\left(\underline{m} \mid \overline{s}\right)}. \tag{19.54}$$

Multiply both sides by $q$ and integrate both sides with respect to $q$ to get

$$E\left[q \mid \underline{m}, \overline{s}\right] = \frac{1}{p\left(\underline{m} \mid \overline{s}\right)} E\left[q \mid \overline{s}\right] - \frac{p\left(\overline{m} \mid \overline{s}\right)}{p\left(\underline{m} \mid \overline{s}\right)} E\left[q \mid \overline{m}, \overline{s}\right]. \tag{19.55}$$

From (19.45), $E\left[q \mid \bar{s}\right] = 0$, hence, we can write (19.55) as:

$$E\left[q \mid \underline{m}, \bar{s}\right] = -\frac{p\left(\overline{m} \mid \bar{s}\right)}{p\left(\underline{m} \mid \bar{s}\right)} E\left[q \mid \overline{m}, \bar{s}\right]. \tag{19.56}$$

On the other hand, in category/situation $\underline{s}$, using (19.45), (19.48) we get,

$$E\left[q \mid \underline{m}, \underline{s}\right] = E\left[q \mid \overline{m}, \underline{s}\right] = E\left[q \mid \underline{s}\right] = 0. \tag{19.57}$$

From (19.56), (19.57), a Bayesian uses the correct model for each category. In category $\bar{s}$, where the signals are informative, the Bayesian forms a different (and correct) posterior relative to category $\underline{s}$. We now describe how a coarse thinker infers quality, conditional on a message $m = \left\{g, b\right\}$.

We use the subscript "$c$" for a coarse thinker. So, $p_c$ refers to the probability assigned by a coarse thinker to some event and $E_c$ refers to the expectation computed by a coarse thinker. In the absence of a subscript, as in $p$, $E$, we refer to the analogous computations of a Bayesian who does not use categorization. We make the following assumption about a coarse thinker.

**A5:** A coarse thinker includes both situations $\underline{s}$ and $\bar{s}$ into the same category, hence, for any signal, $s$,

$$p_c\left(q \mid m, s\right) = p\left(q \mid m, \underline{s}\right) p(\underline{s} \mid m) + p\left(q \mid m, \bar{s}\right) p(\bar{s} \mid m). \tag{19.58}$$

In (19.58), although the coarse thinker is able to observe the situation, the same model is used for both situations. A coarse thinker does not evaluate the informativeness of the message for the particular situation, but rather by the average informativeness across all situations in the category, $s \in \left\{\underline{s}, \bar{s}\right\}$.

Multiply both sides of (19.58) by $q$ and integrate with respect to $q$ to get

$$E_c\left(q \mid m, s\right) = E\left(q \mid m, \underline{s}\right) p(\underline{s} \mid m) + E\left(q \mid m, \bar{s}\right) p(\bar{s} \mid m). \tag{19.59}$$

Since $p(\underline{s} \mid m) + p(\bar{s} \mid m) = 1$ and $p(\underline{s} \mid m) \leq 1$, $p(\bar{s} \mid m) \leq 1$, it follows that the expected quality inferred by the coarse thinker is a linear combination of what a Bayesian would compute. Since $0 = E\left(q \mid m, \underline{s}\right) < E\left(q \mid m, \bar{s}\right)$ it follows from (19.59) that

$$E\left(q \mid m, \underline{s}\right) \leq E_c\left(q \mid m, s\right) \leq E\left(q \mid m, \bar{s}\right). \tag{19.60}$$

This leads to the following straightforward implication.

**Proposition 19.1** *Relative to a Bayesian, the coarse thinker overreacts when the signal is $\underline{s}$ and underreacts when the signal is $\bar{s}$.*

The intuition for the result in Proposition 19.1 is that the coarse thinker acts as a Bayesian who is unable to distinguish between the two situations. When the situation is completely uninformative, he still puts some weight on it being an informative situation, so he overreacts. This can explain the third example above; voters seem to put some weight on Arnold Schwarzenegger's optimism of America's ability to succeed under free trade. They transfer Arnold's ability to win over strong characters in films into another context where such ability has apparently little relevance. Conversely, when the situation is completely informative, the coarse thinker puts some

weight on the situation being completely uninformative, so he underweights the information content of the message.

We now examine the incentives of a persuader to alter the message observed by an individual who might be persuaded. We make the following assumptions.

**A6:** The objective of the persuader is to maximize the expected quality as perceived by the individual. Presumably, an increase in the quality induces a buyer to buy more, or a politician to get more votes.

**A7:** Suppose that the true message is given by $m \in \{\underline{m}, \overline{m}\}$. Denote the message relayed by the persuader to the individual by $m' \in \{\underline{m}, \overline{m}\}$; we allow for $m' \neq m$. The persuader can distort the true message at some cost $c > 0$. The individual is a *face value individual* in the sense that any message given to the individual by the persuader is believed.[19]

In the light of A6, A7, the payoff of the persuader is given by

$$\begin{cases} E(q \mid m', s) & \text{if} \quad m' = m \\ E(q \mid m', s) - c & \text{if} \quad m' \neq m \end{cases}$$

We now examine the incentives of the persuader to distort the message.

**Proposition 19.2** *(Bayesian individual): When $s = \underline{s}$, the persuader always reveals the true message to a face value Bayesian and never incurs a cost of deception, c.*

Proof: From (19.57),

$$E[q \mid \underline{m}, \underline{s}] = E[q \mid \overline{m}, \underline{s}] = 0,$$

hence, by altering a message there is no improvement in average quality. However, the persuader pays a cost, $c$, of altering the message. Hence, it is never optimal to do so. ■

**Proposition 19.3** *(Coarse thinkers): Suppose that $s = \underline{s}$. The persuader might not reveal the true message to a face value coarse thinker. In particular, if c is less than some critical value $c^*$, then the persuader will replace a bad message with a good message.*

Proff: Using (19.57), the first term on the RHS of (19.59) is zero. Furthermore, using Bayes' rule we get that $p(\bar{s} \mid m) = \frac{p(m|\bar{s})p(\bar{s})}{p(m)}$, hence, from (19.58) we get that

$$E_c(q \mid m, s) = E(q \mid m, \bar{s}) \frac{p(m \mid \bar{s})p(\bar{s})}{p(m)}.$$

Using (19.56) we can derive the following expressions for the expected quality for a coarse thinker under the two possible messages $\overline{m}$ and $\underline{m}$.

$$E_c(q \mid \overline{m}, s) = E(q \mid \overline{m}, \bar{s}) \frac{p(\overline{m} \mid \bar{s})p(\bar{s})}{p(\overline{m})}. \tag{19.61}$$

[19] One could alter this assumption to allow for strategic inference by the individual that takes into account the incentives of the persuader to distort the message. This is allowed for by Shleifer et al. (2008), but it does not alter the essential insights.

$$E_c\left(q \mid \underline{m}, s\right) = -E\left[q \mid \overline{m}, \overline{s}\right] \frac{p\left(\overline{m} \mid \overline{s}\right) p(\overline{s})}{p(\underline{m})}. \tag{19.62}$$

For the persuader, replacing the true message $\underline{m}$ with $m' = \overline{m}$ improves profits if

$$\begin{aligned} &E_c\left(q \mid \overline{m}, s\right) - c > E_c\left(q \mid \underline{m}, s\right). \\ \Leftrightarrow\ &c < E_c\left(q \mid \overline{m}, s\right) - E_c\left(q \mid \underline{m}, s\right). \end{aligned} \tag{19.63}$$

Substitute (19.61), (19.62) in (19.63) we get

$$c < c^* \equiv E\left[q \mid \overline{m}, \overline{s}\right] p\left(\overline{m} \mid \overline{s}\right) p(\overline{s}) \left(\frac{1}{p(\overline{m}) p(\underline{m})}\right), \tag{19.64}$$

which completes the proof. ■

We now give some comparative static results.

**Proposition 19.4** *Suppose that $c^*$ is as defined in Proposition 19.3. Then $c^*$ is increasing in (1) the probability with which $s = \overline{s}$ occurs, i.e., $p(\overline{s})$, and (2) the informativeness of message $\overline{m}$ when $s = \overline{s}$, i.e., $E\left[q \mid \overline{m}, \overline{s}\right]$.*

Proof: Directly differentiate both sides of (19.64) successively with respect to $p(\overline{s})$ and $E\left[q \mid \overline{m}, \overline{s}\right]$. ■

From Propositions 19.3, 19.4, an increase in $p(\overline{s})$, $E\left[q \mid \overline{m}, \overline{s}\right]$ increases $c^*$, increasing the likelihood that the persuader lies by giving out good news ($\overline{m}$) when the news is actually bad. The intuition is as follows. From (19.61), (19.62), an increase in $p(\overline{s})$, $E\left[q \mid \overline{m}, \overline{s}\right]$ raises $E_c\left(q \mid \overline{m}, s\right)$ and reduces $E_c\left(q \mid \underline{m}, s\right)$. Thus, the marginal incentive of the persuader to lie and give out the message $\overline{m}$ increases.

DellaVigna and Gentzkow (2010) survey the non-experimental evidence on persuasion; there is much evidence that is supportive of persuasion. They consider four groups of individuals that are typically the target of persuasion: consumers, voters, donors, and investors. They consider questions such as: Are consumers affected by advertising? Is the voting decision of voters influenced by political campaigns? Do newspaper endorsements and endorsements by influential citizens persuade voters to vote differently? Does door to door campaigning for funds persuade donors to give more? Do analyst recommendations influence investment in stocks? The set of questions that falls within the domain of persuasion theories is wide and important. This continues to be an active area of research.

We do not address here the important issues of how and why these categories are formed and how/why they might change over time. These questions lie beyond the scope of the current model, but their resolution will ultimately determine the importance of this class of models. These observations are similar to those we made for *analogy-based equilibria* in Part 4. The issue of optimal categorization is considered in Mohlin (2014), Peski (2011), and Fryer and Jackson (2008). Mengel (2012) considers the evolution of coarse categorization. We close with an example of categorization, which speaks to the issue of racism in the workplace that arises despite the absence of social identity.

**Example 19.5** *(Fryer and Jackson, 2008) Suppose that employees are either black or white. In a population of 100 workers, 90 are white and 10 are black (minority group). Workers' types*

*are either high ability (H) or low ability (L). Half the workers in each race are of each type. Denote a black worker by 0, a white worker by 1, a low ability worker by 0 and a high ability worker by 1. Thus, there are four categories of workers:* $(0,0), (0,1), (1,0), (1,1)$. *Suppose that the employer forms only three categories. Thus, at least one of the categories will contain either two different races, or two different types of workers, or both.*

*Suppose that the employer has previously interacted with a large number of workers in each of the categories, in proportion to their presence in the population. Since there are 10 black workers out of every 100 and half of those are low ability, so the employer must have interacted with 5 workers of type (0,0). Calculating in the same manner, the employer must have interacted with 5 workers of type (0,1), 45 workers of type (1,0) and 45 workers of type (1,1). The mean of any two categories that are grouped together is a population proportions weighted mean. So the mean of the categorization* $\{(0,1), (1,1)\}$ *is* $(0.9, 1)$ *while the mean of* $\{(1,0), (1,1)\}$ *is* $(1, 0.5)$. *The mean of any single category is the category itself; e.g., the mean of* $\{(0,1)\}$ *is just* $(0,1)$. *Let the objective of the employer be to minimize the "sum across categories of the total variation of the mean from each category"; clearly several other objectives are also possible. Fryer and Jackson show that the optimal categorization under such an objective is* $\{(1,0), (1,1), \{(0,0), (0,1)\}\}$.

*Thus, the optimal categorization groups together the two types of black workers in the same category. In this model, blacks are more coarsely sorted not because of racism, or malevolence, but simply because of the underlying objective function and the proportions of blacks in the population. Since the two types of black workers are put into the same category, the representative member of this category is viewed by the employer as* $0.5\,(0,0) + 0.5\,(0,1) = (0, 0.5)$. *This is good news for a low ability black worker but bad news for a high ability black worker. Indeed if the decision to acquire human capital is endogenous, this is likely to reduce the incentive of black workers to acquire high ability.*

## 19.13 Mental models

Under coarse thinking, individuals form categories such that all elements in a category are treated with an identical model. This economizes on the number of mental models required to analyze a large number of events, hence, it economizes on scarce cognitive resources. A useful working definition of *mental models*, based on Denzau and North (1994), is given in WDR (2015, p. 11): "When people think, they generally do not draw on concepts that they have invented themselves. Instead, they use concepts, categories, identities, prototypes, stereotypes, causal narratives, and worldviews drawn from their communities." Chapter 3 in WDR (2015) gives an excellent introduction to mental models. The WDR (2015, p. 11) further writes: "There are mental models for how much to talk to children, what risks to insure, what to save for, what the climate is like, and what causes disease. Many mental models are useful; others are not as useful and may even contribute to the intergenerational transmission of poverty."

Mental models are essential for people to make decisions by economizing on cognitive costs and can be passed down from generation to generation. Mental models can persist even when they are no longer useful, hence, while they economize on cognitive costs, they can lead to suboptimal decisions. An example of harmful mental models is stereotyping of other individuals and racial discrimination based on social group identities; see Chapter 7 and Example 19.5 above. Datta and Mullainathan (2014) document several examples of incorrect mental models that are detrimental to self-interest; let us consider some of these examples.

1. In many developing countries, there is a low take-up of the oral rehydration solution (ORS), a relatively cheap and effective mix of essential salts to treat diarrhea. However, 35% of poor women surveyed in India believed that the best response to diarrhea is to reduce, not increase, fluid intake; they possibly draw an incorrect analogy to a leaky bucket. Since ORS saves lives but does not alleviate the symptoms of diarrhea, it allows many people to hold on to an incorrect, and potentially fatal, mental model.
2. Farmers might overuse nitrogenous fertilizers because they believe that such fertilizers promote greener plants. This model works well for plants like spinach but not as well for grains, because it over-encourages green healthy leafs at the cost of lower crop yields.
3. Parents in Morocco and Madagascar pull their children out of school too early. In their incorrect mental model, education only pays if children go all the way up to secondary school. However, evidence shows that every extra year of schooling is equally valuable. Thus, while parents could afford to send their children to school for longer, doing so is inconsistent with their mental models (Banerjee and Duflo, 2011).

Insofar as undesirable outcomes are not simply the product of cognitive limitations but also of defective mental models, policy must address factors conducive for the formation of appropriate mental models.

Why do "bad" mental models persist? There is more than one potential cause. A range of judgment heuristics that we have already considered in this chapter may provide potential clues. Mental models may serve the purpose of *anchors* and the *availability heuristic* could ensure that existing mental models are the ones most likely to come to mind when taking decisions. *Confirmation bias* could ensure that individuals ignore information that contradicts their existing mental models. Social identity (see Chapter 7) too could play a role. By conforming to preexisting and shared mental models in a social group, individuals may wish to demonstrate that they belong to the group of insiders.

Existing mental models may also foreclose the choices that people could make to test the models (WDR, 2015). For instance, inadequate social exchange between communities that mistrust each other can allow mutual stereotypes to persist. Female genital mutilation may persist in parts of the world because contact with the non-mutilated genitalia may be perceived to be harmful. In the face of this incorrect model, individuals may be unwilling to experiment with alternative mental models, due to the high perceived costs of doing so (Mackie, 1996; Gollaher, 2000).

Although individuals can exhibit heterogeneity in mental models, shared mental models are essential to developing institutions. Indeed, there is a two-directional feedback between mental models and institutions. For instance, in Sierra Leone, the degree of competition for the role of paramount chiefs, who are elected for a lifetime, also determines the quality of governance (Acemoglu et al., 2013). In each area, the degree of competition is fixed because paramount chiefs can only come from families that were originally recognized by the British colonial authorities. Greater competition is found to be conducive to better governance. However, the persistence of the degree of competition in each area has influenced the mental models that people hold for respect towards authority. In areas with lower political competition, people are more respectful of authority and less likely to question the actions of leaders. This, in turn, perpetuates the existing degree of competition.

Some mental models are hardwired in humans due to their evolutionary history (e.g., fear of snakes and spiders; see Chapter 14 for more details). Historical accidents can also explain some mental models. For instance, the current low levels of trust in parts of Africa can be traced back

to the slave trade (Nunn, 2008; Nunn and Wantchekon, 2011). Black Africans required guns to protect themselves against capture by the slave traders. However, because they could not afford to buy guns, they kidnapped and sold other black Africans to slavery, in exchange for money that, in turn, financed guns. This eventually led to mistrust of strangers that still persists in some parts of Africa. In another example, the physical superiority of males gives them a comparative advantage in the use of agricultural implements, such as ploughs. Hence, gender inequities are more pronounced in societies with an agricultural past (Alesina et al., 2013).

Mental models are distinct from norms in that they are not enforced by social norms. Rather, mental models reflect "broad ideas about how the world works and one's place in it" (WDR, 2015).

How does one get people to change incorrect mental models? Datta and Mullainathan (2014) argue that simply telling people that they hold the wrong mental model is unlikely to work. A more promising strategy is to provide evidence that falsifies the core beliefs in an incorrect mental model. For instance, a particularly potent method of getting people to pay attention to reducing a wasteful activity is by comparison with a reference social group or peers. For instance, when individuals are given information about a comparison of their electricity consumption relative to neighbors, they reduce consumption by as much as that caused by an 11–20% increase in the price of electricity (Allcott and Mullainathan 2010; Allcott and Rogers, 2014); a similar effect is documented for water consumption (Ferraro and Price 2013).

Villagers in poor countries are more likely to educate girls if greater information is provided on the availability of jobs for girls with high school degrees (Jensen, 2012). Long-running soap operas in Brazil that emphasized small family size led to reduced fertility, particularly for those of a similar age to the role models in the soap operas (La Ferrara et al., 2012). Exposure to cable TV in India led to a reduction in fertility, increase in women's status, reduction in domestic violence, and an increase in children's school enrollment, suggesting a change in mental models (Jensen and Oster, 2009). Legal or regulatory changes can also induce changes in mental models (Sunstein, 1996); this might have been the case in changes in attitudes towards slavery, wearing helmets, smoking, racism, and casteism.

## 19.14 Herbert Simon's approach to bounded rationality

In neoclassical economics, individuals use subjective expected utility (SEU) to evaluate risky/uncertain situations and employ Bayes' rule to update information. No cognitive limitations are placed on the individual, nor is the application of SEU dependent on the environment that an individual faces. The heuristics and biases approach shows that this view of human nature is often rejected. This has led some to argue that mainstream economic theory is based on an unrealistic picture of human decision making (Selten, 2001).

*Rationality* has several possible interpretations. A common view in economics is that individual behavior is rational if it can be interpreted as the solution to some optimization problem. Of course, the optimization problem must be a plausible one. Otherwise, it is possible to rationalize any behavior by assigning the utility function to take a value 1 for the action chosen by an individual and zero otherwise. The notion of the optimality of behavior, or the quality of the decision, has been termed as *substantive rationality* by Simon (1955). Classical economic theory focuses mainly on substantive rationality and does not give much attention to *procedural rationality*, or the quality of the process of decision.

Substantive rationality did not figure directly in the writings of classical economists. Simon (2000) interprets Voltaire's well-known saying "the best is the enemy of the good" as implying that "if you are too preoccupied with attaining the optimum, you won't even get an acceptable result." Indeed, much of the discourse in economics until well into the middle of the twentieth century used the notion of a *reasonable person* rather than that of a *utility maximizing rational person*.

Is the rationality hypothesis compelling on grounds of empirical evidence? The evidence is mixed at best, and success is mixed with well-documented and prominent failures. Börgers (1996) concludes that "Unfortunately, there are no striking examples of empirical regularities, which are well explained by the rationality hypothesis." Substantive rationality, as used in neoclassical economics, prescribes choosing on the basis of marginalist conditions; for instance, firms produce at the point where marginal revenues equal marginal costs. In his Nobel lecture, Simon (1978, p. 347) is very explicit on this issue: "But there are no direct observations that individuals or firms do actually equate marginal costs and revenues."

Furthermore, Selten (2001) writes. "However, there is overwhelming experimental evidence for substantial deviations for Bayesian rationality (Kahneman et al. 1982): people do not obey Bayes' rule, their probability judgements fail to satisfy basic requirements like monotonicity with respect to set inclusion, and they do not have consistent preferences, even in situations involving no risk or uncertainty." A more positive view of rationality is taken by Gintis (2009, p. 2) who argues for maintaining at least a minimum degree of rationality based on *preference consistency*; he argues that violations of preference consistency can be accounted for by changes in context, frames, and reference points.

The impact of the rationality assumption on other social sciences has not been profound. In evaluating the impact of such work, with particular reference to the collected works of Becker (1981), Simon (2000) says the following. "Perhaps its greatest failing was in not producing major insights into political and social processes that were not already well-known, if in somewhat less qualitative and less formalized statements."

In response to these concerns Herbert Simon proposed *bounded rationality* as an alternative approach. The term bounded rationality has become fashionable in a variety of contexts in economics and its usage and meaning have become variegated. To be specific, we follow the original usage of the term in Simon (1956, 1957). Simon used the metaphor of the two blades of a pair of scissors. One blade represents *cognitive limitations* of humans, while the second blade emphasizes the *structure of the environment*. To cut successfully, the scissors must use both blades, hence, for a proper approach to bounded rationality, one cannot choose to omit either cognitive limitations or the structure of the environment. By contrast, the practice in most of economics is to assume unbounded cognitive powers and underemphasize the structure of the environment, such as framing effects, context, and emotions.

Bounded rationality does not imply that individuals are irrational. Hence, bounded rationality is not synonymous with erratic human behavior. Boundedly rational individuals often solve some well-specified optimization problem. However, whether individuals should be modeled as optimizing agents is an open question. For arguments on both sides, see the "Forum on the Role of Bounded Rationality versus Behavioral Optimization in Economic Models" published in the June 2013 issue of the *Journal of Economic Literature*. The main difference between bounded rationality and substantive rationality (as used in neoclassical economics) is that, in the former, the optimization problem respects the cognitive bounds of the relevant economic actors and the structure of the environment. Bounded rationality is, thus, often concerned with *procedural rationality*. It tries to open up the black box of human decision making, and the steps that humans take to formulate and solve problems in different environments.

Bounded rationality has sometimes been interpreted as *optimization under constraints* because the information necessary for taking actions involves the costs of search. In one view, if one introduces constraints that capture the costs and benefits of search, then one could take account of bounded rationality (Stigler, 1961). However, the problem with this argument is that to estimate the costs and benefits of search, one would have to first estimate the costs and benefits of estimating the costs and benefits of search. One can construct an infinite regress of such search, which can be even more informationally demanding than the original problem. This compounds the very problem that these additional constraints were designed to address and makes the model even more psychologically implausible.

To see this slightly more formally, we can use the following intuitive construction in Selten (2001). He distinguishes between *familiar* and *unfamiliar* problems. For familiar problems, the decision maker *knows* the set of alternatives (or trivially knows where to find them) and also knows the *method* for choosing among these alternatives. For unfamiliar problems, the set of alternatives and methods for choosing among them are both unknown. Thus, in this case, the decision making process can be understood as being composed of the following two levels.

1. (Level 1) Find the set of alternatives to be chosen.
2. (Level 2) Find a method of making a choice among the set of alternatives.

A decision maker who is ignorant of what to do at Level 1, will also not know what to do at Level 2. Thus, we get:

3. (Level 3) Find an optimal method of solving the problem at Level 2.

Hence we get an infinite regress, with level $k$ being:

$k$. (Level $k$) Find the optimal method for solving Level $k-1$.

While this is not a formal proof, it is strongly suggestive of the impossibility of classical optimization in limited time. In the words of Selten (2001) "Trying to optimize in such situations is like trying to build a computer that must be used in order to determine its own design."

Simon (1978) in his Noble lecture is quite skeptical about theories that explicitly incorporate costly search and information. He offers the following sobering views (pages 358–9): "In none of these theories . . . is the assumption of perfect maximization abandoned. . . . Hence, the new theories do nothing to alleviate the computational complexities facing the decision maker . . . but simply magnify and multiply them. . . . Hence, to some extent, the impression that these new theories deal with the hitherto ignored phenomenon of uncertainty and information transmission is illusory. For many economists, however, the illusion has been pervasive."

### 19.14.1 *Aspiration adaptation theory*

In an important class of bounded rationality models, the individual, or an institutional entity such as a firm, has a goal or a reference level, referred to as an *aspiration level* (Simon, 1957). Aspiration levels may refer, for instance, to the target profit of a firm, a target income for a consumer, or even a level of employment/growth rate for a government. A decision alternative is *satisfactory* if it meets or exceeds the aspiration level.

Decision alternatives are not exogenous but must be searched by the individual. To avoid an infinite regress problem, the decision maker can use a simple stopping rule. For instance, stop when an alternative (or the outcome from an alternative) exceeds the aspiration level. Such an

approach has also been called *satisficing* to distinguish it from *optimization*. The aspiration levels are not fixed, but evolve over time.

One simple method of updating aspiration levels is to adjust them downwards if successive alternatives fall below the aspiration level, much as one would reduce the sale price of a house that has been on the market for too long. As Simon (1978) puts it: "In a benign environment that provides many good alternatives, aspirations rise; in a harsher environment, they fall." The individual has a well-defined optimization problem. Unlike the models in neoclassical economics that are based on substantive rationality, this model incorporates: (1) Limited cognitive abilities that are reflected in a simple stopping rule, and (2) the structure of the environment, as embodied in the formation of aspirations.

An early attempt at a formal model of *aspiration adaptation* was by Saurmann and Selten in 1962; an English version was published as Selten (1998). Consider the role played by aspirations in a symmetric two-player normal form stage game, denoted by $G(2, \{C, D\}, \{C, D\}, u_1, u_2)$, that is repeated over discrete time, $t = 0, 1, 2, \ldots$.(Cho and Matsui, 2005). The two players are indexed by $i = 1, 2$, and player $i$ has the strategy set $S_i = \{C, D\}$; think of these actions as, respectively, *cooperation* (C) and *defection* (D).

Denote the action of player $i$ in period $t$ by $s_{it} \in S_i = \{C, D\}$ and denote the action profile in any period $t$ by $s_t = (s_{1t}, s_{2t})$. The respective payoffs of players 1, 2 in the stage game are denoted by $u_1$ and $u_2$ such that $u_i : \{C, D\} \times \{C, D\} \to \mathbb{R}$. We assume that players are sufficiently patient so that if both players were rational in the classical sense, then mutual cooperation $(C, C)$ can be sustained as an equilibrium of an infinitely repeated game. However, we assume that players are boundedly rational.

In the spirit of Herbert Simon's work, suppose that players are bounded rational and form aspiration levels. Since players do not have infinite cognitive abilities, it is assumed that they summarize the history of the game into a single number, the *aspiration level*, which is simply an average of past payoffs of the player; the precise sequence of outcomes and actions in the past is not relevant. Hence, the aspiration level of player $i$ at time $t$, $a_{it}$, is given by

$$a_{it} = \frac{1}{t} \sum\nolimits_{\tau=1}^{t} u_i (s_\tau), t = 1, 2, \ldots \tag{19.65}$$

Rewriting (19.65), we get

$$a_{it} = a_{it-1} + \frac{1}{t} [u_i (s_t) - a_{it-1}]. \tag{19.66}$$

From (19.66), the aspiration level evolves in a manner that is reminiscent of *adaptive expectations*. Current aspirations are equal to lagged aspirations adjusted for the difference between the current actual payoff and lagged aspirations. Initial aspiration levels of the two players are exogenously given by $a_0 = (a_{10}, a_{20})$, and initial actions are also exogenously given by $s_0 = (s_{10}, s_{20})$.

The significance of the aspiration level $a_{it}$ is that it is the payoff which the individual expects from the action $s_{it}$ in the current period, conditional on the past payoffs. So, if past payoffs were high (low), then current aspirations are also high (low). Since player $i = 1, 2$ has only two actions, we denote the chosen action at time $t$ by $s_{it}^{+} \in S_i$ and the unchosen action by $s_{it}^{-} \in S_i$; for example, if the chosen action is $C$, then $s_{it}^{+} = C$ and $s_{it}^{-} = D$. The individual uses the following *behavioral rule* to choose actions

| | C | D |
|---|---|---|
| C | 1, 1 | −1, −1 |
| D | −1, −1 | 0, 0 |

**Figure 19.7** A coordination game to study the role of aspiration adaptation.

Source: Reprinted from *Journal of Economic Theory* 124(2): 171–201. Cho, In-Koo, and Akihiko Matsui (2005). "Learning aspiration in repeated games." Copyright © 2005, with permission from Elsevier.

$$s_{it+1} = \begin{cases} s_{it}^{+} & \text{if} \quad u_i \geq a_{it} \\ ps_{it}^{+} + (1-p)s_{it}^{-} & \text{if} \quad u_i < a_{it} \end{cases}, \tag{19.67}$$

where $0 < p < 1$ and $ps_{it}^{+} + (1-p)s_{it}^{-}$ is a mixed strategy that plays $s_{it}^{+}$ with probability $p$ and $s_{it}^{-}$ with probability $1-p$. The idea behind the behavioral rule is simple. If the current payoff from an action is at least as good as the current aspiration level ($u_i \geq a_{it}$), then continue playing the current action, $s_{it}^{+}$. Otherwise, switch to the other action, $s_{it}^{-}$, with some probability $1-p$. Thus, $p$ represents a measure of *inertia*. If $p = 0$, then the individual completely switches to the other action (no inertia) and if $p = 1$, then the individual never switches (full inertia). In order to make the problem interesting, we rule out both possibilities by assuming $0 < p < 1$. One possible cause of inertia is potential switching costs that we ignore for pedagogical reasons.

Given the initial aspiration levels, $a_0$, and initial actions, $s_0$, we can use the definition of aspirations in (19.66) and the behavioral rule in (19.67) to generate a time profile of aspiration levels and actions $\{a_\tau, s_\tau\}_{\tau=1}^{\infty}$.

Consider the coordination game in Figure 19.7. If both players coordinate at $(C, C)$, they get the highest payoff of 1 each. Lack of coordination pushes them to the lowest possible payoff tuple $(-1, -1)$. However, if they both do not cooperate, then both get a zero payoff, corresponding to the strategy profile $(D, D)$. Assume that the initial aspiration levels are given by

$$a_{10} = a_{20} = -0.1. \tag{19.68}$$

1. Suppose that the initial strategy profile is $s_0 = (C, C)$. Then, the payoffs of the two players at time $t = 0$ are given by

$$u_1(C, C) = u_2(C, C) = 1. \tag{19.69}$$

Since $u_i(C, C) > a_{i0}$, (19.67) implies that the strategy profile at $t = 1$ is $s_1 = (C, C)$. From (19.66), the aspiration levels in period $t = 1$ are given by $a_{i1} = a_{i0} + [u_i(C, C) - a_{i0}]$. Using (19.68), (19.69), $a_{i1} = 1$. Since $u_i(C, C) = a_{i1}$ in period 1, it follows from (19.67) that in $t = 2$, the equilibrium action profile is again $(C, C)$. The aspiration level in period 2 is $a_{i2} = a_{i1} + \frac{1}{2}[u_i(C, C) - a_{i1}] = 1$. Repeating the usual steps, at $t = 3$, the action profile is again $(C, C)$. Extending this argument, the prediction of the model is that if both players start by coordinating, they continue coordinating. The reason is that the payoff from $(C, C)$ is always at least as good as the aspiration level, hence, the players never feel the need to experiment and choose a different action.

2. Suppose that the action profile $(C, D)$ is played in the initial period; the game is symmetric, so the results with the profile $(D, C)$ are identical. The payoffs of the two players at time $t = 0$ are given by

$$u_1(C, D) = u_2(C, D) = -1. \tag{19.70}$$

From (19.68), (19.70), $u_i(C, D) < a_{i0}$, so (19.67) implies that at time $t = 1$, each player switches to the alternative action with probability $1 - p$. Thus, the actions chosen by the two players at $t = 1$ are given by

$$s_{11} = pC + (1 - p)D. \tag{19.71}$$

$$s_{21} = pD + (1 - p)C. \tag{19.72}$$

We can now compute the payoffs of the two players in period 1 using the mixed strategies of the two players given in (19.71), (19.72). These can be calculated to be

$$u_1(s_{11}, s_{21}) = u_2(s_{11}, s_{21}) = (1 - p)(2p - 1) - p^2. \tag{19.73}$$

The payoffs of both players are equal because of the symmetry of the game.
Consider the polar case of no inertia, i.e., $p = 0$. Then at $t = 1$, using (19.71), (19.72), both players switch their actions, so that the time $t = 1$ action profile is $(D, C)$. Evaluating (19.73) at $p = 0$, we have $u_1 = u_2 = -1$. When $p = 0$, the aspiration levels of the two players at date $t = 1$ are given by $a_{11} = a_{10} + [u_1(D, C) - a_{10}]$ and $a_{21} = a_{20} + [u_2(D, C) - a_{20}]$. Using (19.68), (19.70),

$$a_{11} = a_{21} = -1. \tag{19.74}$$

Since $u_1 = a_{11}$ and $u_2 = a_{21}$, the behavioral rule in (19.67) implies that the actions at $t = 2$ are the same as those at $t = 1$, so the action profile remains $(D, C)$. A simple calculation shows that $a_{12} = a_{22} = -1$, so aspirations are not revised in the next period and payoffs do not change either. Thus, given an initial action profile $(C, D)$, if there is no inertia, the game settles at the profile $(D, C)$ for all time periods, starting at $t = 1$. Analogously, one can show that, with $p = 0$, starting from $(D, C)$, the game settles at $(C, D)$ from $t = 1$ onwards. Now consider the case when there is inertia, $p > 0$. In this case, from (19.66) and (19.73),

$$a_{11} = u_1 > -1, a_{21} = u_2 > -1 \tag{19.75}$$

for all $p \in (0, 1)$. The behavioral rule (19.67) implies that actions are not altered in the next period ($t = 2$), so $s_{12}$ and $s_{22}$ are given, respectively, by the mixed strategy profile in (19.71), (19.72). Utilities continue to be given by (19.73) in the second period also. The aspiration levels in the second period for $i = 1, 2$ are $a_{i2} = a_{i1} + \frac{1}{2}(u_i - a_{i1})$, where $a_{i1}$ is given by (19.75), thus, $a_{i2} = a_{i1}$. Clearly, the mixed strategies in (19.71), (19.72) are played every period from $t = 1$ onwards. Identical results follow in the case where the initial action profile is $(D, C)$.

3. Suppose that the initial action profile is $(D, D)$. The first period payoffs are then $(0, 0)$. Since $0 > a_{i0} = -0.1$, both players' aspiration levels are met and they stick to the profile $(D, D)$

in period 1. One can use this argument repeatedly to show that the action profile $(D, D)$ is followed in every period. The aspiration level improves in every period and converges eventually to zero, from below.

Based on this analysis, the following observations can be made.

The equilibrium of the game is heavily dependent on the initial state of play. Furthermore, the cooperative outcome can be sustained in an equilibrium of the aspirations adaptation game. For the parameter values chosen in the example, this outcome arises if players begin by cooperating in the initial period. Non-cooperation or defection, however, begets further rounds of non-cooperation. Finally, players might coordinate on the defection strategy and never cooperate again.

In a more general model, Cho and Matsui (2005) introduce other extensions. They introduce a small probability of players playing a different action; one interpretation is that players make a mistake with some positive, but small, probability. It turns out that the inefficient coordination outcome $(D, D)$ is not robust to this perturbation and the only equilibrium which survives is the "good" coordination equilibrium.

It should be quite intuitive that this framework can be used to support the cooperative outcome in a range of games (Colman et al., 2010). For instance, in the prisoner's dilemma game, if aspiration levels are sufficiently high, then the non-cooperative outcome results in payoffs that are lower than the aspiration level. Individuals then experiment with other actions (such as cooperation). If both players happen to cooperate at the same time, then the payoffs exceed the aspiration level, in which case there is no further need to experiment with other actions. These ideas can be formalized in different ways. Börgers and Sarin (2000) and Karandikar et al. (1998) are two such attempts. Both papers contain the idea that aspirations are updated slowly in the direction of the eventual cooperative payoffs. Karandikar et al. (1998), in addition, introduce a small probability of trembles in actions and demonstrate in $2 \times 2$ games that most players will cooperate most of the time.

Aspirations in these models are based on plausible adaptive processes. However, ultimately empirical evidence will determine the precise nature of the formation of aspirations, which might well be context and situation dependent. In one interesting experiment, Bernard and Taffesse (2014) randomly selected households in Ethiopia to watch inspirational videos based on the actual life stories of people who had successfully followed goals and achieved a higher socio-economic position. Six months later, these individuals had higher savings rates and higher aspirations, particularly about their childrens' educational future. It is also well known that poverty depresses aspirations and engenders a feeling of greater hopelessness (Ray, 2006; Haushofer and Fehr, 2014). Similar results are found in richer countries where poorer students have relatively lower aspirations towards education and employment (Guyon and Huillery, 2014).

In the models above, there is a single goal, feedback is perfect, and the underlying economic environment is commonly understood. In actual practice, however, individuals, institutions, and governments, often try to achieve complex long-term goals. For instance, (i) individuals might be interested in long-term happiness, (ii) firms in long-term profits, and (iii) governments in long-term economic prosperity. Yet it might not be clear how to achieve these complex goals. On the other hand, there could be a set of short-term goals that the individual believes are correlated with long-term goals. Corresponding to the three long-term goals mentioned above, examples of relevant short-term goals could be: the individual practices daily meditation, the firm invests in short-term improvements in customer relations, and the government appoints a particularly capable central banker.

Selten et al. (2012) outline a formal model of the decisions taken by a monopolist, which has these features; an approach that they call *goal systems*. In this framework, economic entities face a complex problem whose solution is unlikely to be obtained by purely rational analysis, although a simulated solution is available by using specialized optimization software. Economic entities then choose among a set of goal variables that are essentially short-term feedback variables. For instance, the monopolist chooses short-term quality of the product and advertisement expenditure that contributes towards raising demand for a product in that period. For each goal, they set an aspiration level that they can adapt upwards or downwards. If no upward adjustments in aspiration levels are feasible, then no further attempts at adjusting aspiration levels is made. Their experiments confirm some of the basic predictions of the aspiration adaptation model.

## 19.14.2 *Fast and frugal heuristics*

Humans do not possess unlimited cognitive abilities, nor do they have unlimited time to solve problems. Thus, as Herbert Simon suggested, modeling *procedural rationality*, i.e., the cognitive process and the decision environment in which decisions are made is important.

Rubinstein (1998) provides several examples of procedural rationality. In chapter 11 of his book, titled *Final Thoughts* he provides a summary of his correspondence with Herbert Simon regarding his book. Simon objects to the lack of empirical support for the rationality procedures introduced by Rubinstein. Simon is quoted as saying "Facts do not come from the armchair, but from careful observation and experimentation" Rubinstein (1998, p. 188). Rubinstein's own response is detailed in his chapter 11, but the following remark illustrates well his position. "The crowning point of making microeconomic models is the discovery of simple and striking connections between concepts (and assertions) that initially appear remote" (Rubinstein, 1988, p. 191). Clearly, Rubinstein's approach, while important and timely in terms of its proposed formalism, is not in the same tradition as Herbert Simon's. Indeed, there is a basic disagreement about the method and purpose of economic models; this book's view has already been laid out in the introductory chapter.

An approach that is closer to Herbert Simon's own views can be found in the work of Reinhard Selten and Gerd Gigerenzer and colleagues at the ABC Research Group, based at the Max Planck Institute of Human Development. Central to the work of this group is the intriguing metaphor that humans draw upon an *adaptive toolbox of heuristics*, conditional on the environment that they find themselves in. Two excellent summaries of this approach are Gigerenzer et al. (1999) and Gigerenzer and Selten (2001). This approach differs fundamentally from the heuristics and biases approach of Kahneman and Tversky.

The view taken by Gigerenzer and colleagues is, in many respects, similar to the *Enlightenment view*, associated with the names of Laplace, Boole, and Piaget; see Gigerenzer et al. (1999) for a representative position. In this view, the normative and descriptive models of human inference are the same as the relevant statistical models. So, for instance, in updating probabilities, the expectation is that humans will use the Bayes' rule.

On the other hand, the work of Kahneman, Tversky, and others (see above) has shown a divorce between the normative view of humans and actual behavior (Kahneman and Tversky, 2000). So, for instance, the behavior of individuals might not conform to the predictions of subjective expected utility theory (SEU), or to the predictions of classical statistics. In his Nobel lecture, Simon (1978, p. 362) is very explicit about the implications: "On the basis of these and other pieces of evidence, the conclusion seems unavoidable that the SEU theory does not provide a good prediction—not even a good approximation—of actual behavior."

Gigerenzer and colleagues take the position that human heuristics are not a *bad* but a *good*. They argue that these heuristics are the appropriate response by humans, with limited cognition and time, to their environment. This is reflected in the title of one of their books: *Simple Heuristics That Make Us Smart*. They argue that the heuristics used by humans can outperform other statistical approaches to solving problems. Gigerenzer and colleagues also challenge the Kahneman–Tversky inspired biases on several other grounds. For instance, they argue that if the problems are presented to subjects in a natural frequency format (as opposed to probabilities) and in a knowledge domain familiar to them, then the biases are significantly reduced or eliminated. We have already examined the evidence on some of these positions in the sections above; we postpone a further discussion of these issues to Section 19.15.

## TWO EXAMPLES OF FAST AND FRUGAL HEURISTICS

We start with two examples that illustrate fast and frugal heuristics. The perceptive reader will note that there is no necessary conflict in the insights from these problems with the Kahneman–Tversky approach, an issue that we return to in Section 19.15.

**Example 19.6** *(The problem of catching a ball): Consider, as in Figure 19.8, the problem of catching a ball in cricket, baseball, or football (by a fielder in the first two cases and by the goalkeeper in the third). Suppose that the origin of the ball is at point A. It travels along the thick parabola and lands at point B. A catcher is stationed at point C. How does he catch the ball? The optimizing, fully rational, approach in economics would suggest the following algorithm.*

*Use the two-parameter equation of a parabola. Take the point of origin of the ball and the initial velocity as the initial conditions of the problem. In order to determine the two parameters of the parabola, take account of factors such as the wind speed, humidity, and spin on the ball. Also account for the fact that these conditions could create perturbations from the shape of an ideal parabola. Then design a machine (unless one is handy) that takes account of all this information in order to compute point B. Since the shortest distance between two points is a straight line, the prediction of the rational model is that the catcher runs in a straight line from C to B to successfully catch the ball.*

*Of course, a weaker position, and the one typically taken in neoclassical economics, is that the calculation process need not be taken literally. Individuals are expected to behave "as if" they can perform the appropriate calculations.*

*In either case, it is very easy to test this prediction. When the runs of successful catchers are observed, it turns out that they do not run in a straight line from C to B (McLeod and Dienes, 1996). Rather, they start running as soon as the ball leaves point A and take the curved, dotted, path shown in Figure 19.8. Hence, the prediction of the standard model is easily rejected in the strict and the weak interpretations.*

*McLeod et al. (2006, p. 139) describe why catchers behave in this manner: "the path of a fielder running to catch a ball is determined by the attempt to satisfy 2 independent constraints. The 1st is to keep the angle of elevation of gaze to the ball increasing at a decreasing rate. The 2nd is to control the rate of horizontal rotation necessary to maintain fixation on the ball." The important feature of this heuristic is that it does not use all the available information. Hence, it is frugal in the use of information. Because it uses only two simple constraints, the heuristic is incredibly fast (the catcher starts running immediately).*

**Example 19.7** *(Dealing with heart attack patients): Suppose that the emergency unit of a hospital receives a patient with chest pains and a suspected heart attack. How best should*

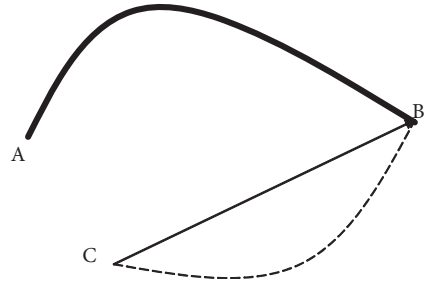


**Figure 19.8** The problem of catching a ball.

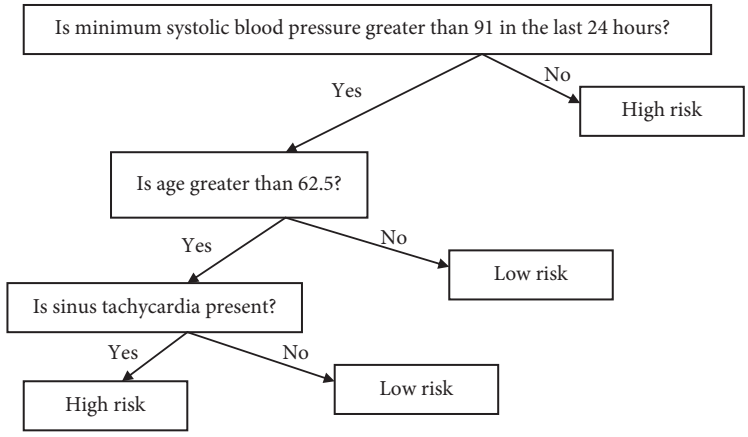


**Figure 19.9** Classifying the degree of risk of suspected heart attack patients.

Source: Peter M. Todd and Gerd Gigerenzer (2000). "Précis of simple heuristics that make us smart." *Behavioral and Brain Sciences* 23: 727–41.

*they deal with the patient? One possibility is to measure all the possible indicators of a heart attack. Then use some statistical procedure to aggregate this information into a number; when this number exceeds some critical value it is almost sure to correctly diagnose a heart attack. However, time, information and cognitive abilities are limited.*

*Breiman et al. (1993) suggest a simple decision tree, which is able to classify patients by the degree of risk that they face; see Figure 19.9. Only three variables are used. If the systolic blood pressure is below 91, the patient is immediately classified as high-risk and no further information is needed. If not, then patients older than 62.5 years and suffering from sinus tachycardia are classified as high risk. Otherwise they are low risk.*

*The distinguishing feature of this simple decision tree, which really is a heuristic to identify high-risk patients, is that it does not try to use all possible information that could have been used. Furthermore, it is simple in the following two respects. It uses qualitative information with binary answers to decision criteria. So, for instance, the absolute difference in one's actual age from 62.5 years is irrelevant; only the cutoff value of 62.5 years is important. The heuristic does not attempt to combine the three variables into a single index. Hence, this heuristic is also fast and frugal. But does it work? Breiman et al. (1993) show that it is actually more accurate*

*than some other complex statistical procedures that are currently used. Hence, fast and frugal heuristics are not necessarily inaccurate or inefficient, but they can be in other cases.*

The main difference between *satisficing* in the sense of Herbert Simon, and *fast and frugal* heuristics is this: satisficing need not always be a fast and frugal heuristic. Some forms of satisficing might even be computationally challenging, while fast and frugal heuristics never are.

A distinguishing feature of the work on fast and frugal heuristics has been the attempt to (1) pit the predictions of such models against the data, and (2) test the efficiency of such heuristics in comparison with standard statistical procedures. As an illustration of this approach, we first consider some examples from Gigerenzer and Goldstein (1996).

### THE RECOGNITION HEURISTIC AND THE TAKE-THE-BEST HEURISTIC

In this section we describe a prominent example of a fast and frugal heuristic. Gigerenzer and Goldstein (1996) ask how individuals might solve the problem of comparing the relative size of different German cities? For instance, which city has a higher population? (a) Hamburg (b) Cologne. They suggest the following procedure that uses limited knowledge and limited cognitive abilities to make a relatively fast decision.[20]

Consider a *reference class* $R$ of cities; this is just a list of cities. For $a, b \in R$ we ask the individual: which is bigger? In answering this question, individuals tap into some *knowledge base*. Essentially, a knowledge base, $C$, consists of *probability cues* $C_0, C_1, C_2, \ldots, C_n$. We focus, here, on cues that take a binary value. For instance, the cue $C_i$ could be: does the city have a football team? The answer is either yes, no, or don't know; we denote these three cases, respectively, by $+, -, ?$. As an example, suppose that we have four reference cities and five cues, so $R = \{a, b, c, d\}$ and $C = \{C_0, C_1, \ldots, C_5\}$. Let the first cue, $C_0$, be the *recognition cue*, i.e., does one recognize a city? The individual can either recognize $(+)$, not recognize $(-)$, or does not know if he can recognize $(?)$. The cue values for each of the four cities are shown in Table 19.9.

The table illustrates two features. First, it uses *limited information* (in actual practice many other cues could have been used). Second, it shows *limited knowledge*. This is shown by the large number of ? signs, which indicate that the cue value is uninformative. Notice that for city $d$, the recognition cue indicates that it has never been heard of by the individual. For this reason, for all other possible cues, $d$ takes a value ?. Let $C_i(x)$ denote the cue value taken by the $i$th cue when the city is $x \in R = \{a, b, c, d\}$. Let $t(x)$ be the question asked from the individual or the *target variable*; in this case this is the population of city $x$. Suppose that for any two cities $x, y \in R$ and

**Table 19.9** The cue values for each of the four cities.

| | $a$ | $b$ | $c$ | $d$ |
|---|---|---|---|---|
| $C_0$(Recognition) | $+$ | $+$ | $+$ | $-$ |
| $C_1$ | $+$ | $-$ | ? | ? |
| $C_2$ | ? | $+$ | $-$ | ? |
| $C_3$ | $-$ | $+$ | ? | ? |
| $C_4$ | ? | $-$ | $-$ | ? |
| $C_5$ | ? | ? | $-$ | ? |

[20] This draws on earlier work on probabilistic mental models; for instance, Gigerenzer (1993), Gigerenzer et al. (1991).

using cue $C_i$ we get that $C_i(x) = +$ and $C_i(y) = -$. Then we say that cue $C_i$ is able to predict correctly which city has the larger population, i.e.,

$$C_i(x) = +, C_i(y) = - \Rightarrow t(x) > t(y). \tag{19.76}$$

Now suppose that over all $x, y \in R$ we compute the number of cases, $N$, where (19.76) holds. Let $M$ be the number of cases where it does not hold, i.e., where, for instance, $C_i(x) = C_i(y) = +$ and $C_i(x) = C_i(y) = -$. Ignore the cases where a cue takes a value ? ("don't know") because no comparison can be made.

Two features of cues, *ecological validity* and *discrimination rates*, are important. These are defined below.

**Definition 19.4** *(Ecological validity): The ecological validity $v_i$ of cue $C_i$ is the frequency with which it predicts the target from among the choices in the reference set, i.e.,*

$$v_i = \frac{N}{N+M}.$$

*So, for instance, let R be the set of German cities with a population of at least 100,000. Let cue $C_i$ check if the city has a football team. In 87% of the cases, the city with the football team has greater population. Thus, cue $C_i$ has an ecological validity of 0.87. Applying similar calculations to all cues we can rank the cues in term of their ecological validity.*

Ecological validity reflects the phenomenon of limited information but not limited knowledge; it does not take account of the ? values. Now suppose that there are a total of $\overline{N}$ pairs of cities $x, y \in R$ such that $x, y$ are distinct. Also let $\hat{N}$ be the number of cases (among all $x, y \in R$) such that the cue is able to discriminate between two options, i.e., either $C_i(x) = +, C_i(y) = -$ or $C_i(x) = -, C_i(y) = +$.

**Definition 19.5** *(Discrimination rates): The discrimination rate, $d_i$, associated with some cue $C_i$ is the relative frequency with which a cue discriminates among two objects in the reference class. Hence,*

$$d_i = \frac{\hat{N}}{\overline{N}}.$$

As $d_i$ increases, the *inferential usefulness* of the cue, $C_i$, increases because it is able to distinguish among a larger number of cases.

Reverting back to the data provided in Table 19.9, how will an individual make a choice? One possibility is for individuals to use the take-the-best algorithm. In this algorithm, the cues are ranked in terms of their ecological validity first.[21] Following the recognition cue, the remaining cues are arranged in the following order. The cue with the highest validity ($C_1$) at the top (also called the best cue) and the one with the lowest validity ($C_5$) at the bottom. The basic principle of this algorithm is to take-the-best and "ignore the rest." The steps followed in the take-the-best algorithm are shown in Figure 19.10.

[21] Another heuristic, the *minimalist heuristic*, ranks cues in a completely random order.

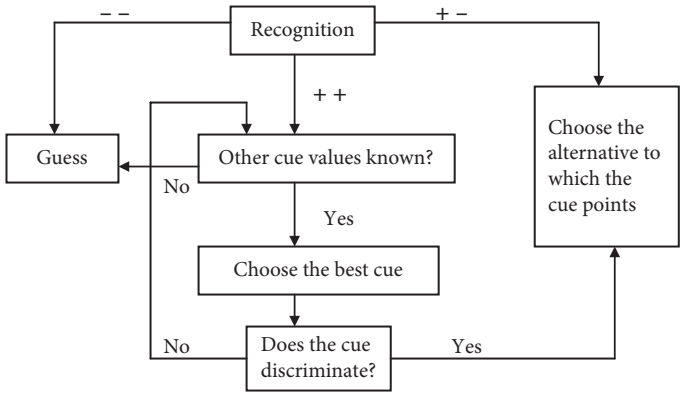


**Figure 19.10** Description of the take-the-best algorithm.

Step I. The first cue that is used is the recognition heuristic. In making a choice among two objects, if one of the objects is recognized and the other is not $(+-)$, then choose the recognized object. If both objects are unrecognized $(--)$, then guess. If both are recognized $(++)$, then go to step II.

Step II. Use the cue with the highest ecological validity. If this cue discriminates among the two options $x, y \in R$ (i.e., $C_i(x) = +, C_i(y) = -$ or $C_i(x) = -, C_i(y) = +$ or $C_i(x) = +, C_i(y) = ?$ or $C_i(x) = ?, C_i(y) = +$), then choose the option that has a value $+$ for the cue. Otherwise, if the cue is unable to discriminate, examine the cue with the next highest ecological validity and repeat step II.

Step III. If one runs out of cues and none is able to discriminate between the two options, then simply guess.

These steps, the definitions of ecological validity, and Table 19.9 can be used to illustrate several examples.

**Example 19.8** *Which of the two cities, a, d, has more inhabitants? Using Step I, the recognition cue, city a is recognized while city d is not, hence, the choice is made at the first step of the algorithm: city a is chosen.*

**Example 19.9** *Which of the two cities, a, b, has more inhabitants? Using Step I, the recognition cue is unable to discriminate between the two cities. Using Step II, we now move on to the cue with the next highest validity, i.e., cue* $C_1$.[22] *It immediately distinguishes between cities, a, b because* $C_1(a) = +, C_1(b) = -$. *Hence, city a is chosen.*

**Example 19.10** *Which of the two cities, b, c, has more inhabitants? In Step I, the recognition cue is unable to distinguish between the two (both are recognized). Cue* $C_1$, *with the next highest validity, is also unable to distinguish because the cue value for city c is unknown* $(C_1(b) = -, C_1(c) = ?)$. *The cue with the next highest validity,* $C_2$, *is able to distinguish* $(C_2(a) = +, C_2(b) = -)$, *and on that basis, city b is chosen.*

These examples share some common features. First, only a limited part of the total stock of knowledge is searched. Second, the stopping rule is simple: "stop when the first discriminatory

[22] Cues $C_1$, $C_2$, and $C_3$ have identical ecological validity in this example.

cue is found." Third, no attempt is made to integrate all the available information, using classical statistical procedures. These are all distinguishing features of fast and frugal heuristics. The take-the-best algorithm is sometimes said to be *non-compensatory* because the discriminating cue with the highest validity determines the final choice; no combination of the other cues can overturn this choice.

The take-the-best algorithm has been used successfully in several contexts. These include: the *confidence frequency effect*, the *less is more effect*, and why the confirmation bias explanation of the overconfidence effect may not be supported by the evidence.[23]

In order to test the take-the-best algorithm in the context of city size, Gigerenzer and Goldstein (1996) consider 83 German cities with at least 100,000 inhabitants. Population is the target variable; 9 binary ecological cues are used, so there are $9 \times 83$ cue values. These cues included the following: Is the city a national capital? Does the city have a football team in the major league? Is the city on the intercity train route? Is the city a state capital? Is the city home to a university? Is the city in the industrial belt? The ecological validity of the cues is decreasing in the order written above.

The point of the exercise is to compare the take-the-best heuristic with some other statistical procedures, all of which use more information, more time, and more computation power. The statistical procedures used include tallying, weighted tallying,[24] unit weight linear model, weighted linear model,[25] and multiple regression. The alternatives against which these heuristics are compared are chosen by the experimenter. There need not be universal agreement that these

[23] For the *confidence frequency effect*, see Gigerenzer et al. (1991); the original take-the-best algorithm is also presented in this paper. For the *confirmation bias explanation of the overconfidence effect*, see Gigerenzer et al. (1991).

[24] Under tallying, first arrange the cues in decreasing order of validity. The recognition cue is not used. The idea is to take account of the (positive) weight of the evidence for a decision. So, for instance, in comparing cities $a, b$, tally the positive cue values for each city across all cues. Then choose the city with the largest number of positive cue values. Thus, city $a$ is chosen over city $b$ if

$$\sum_{i=1}^{n} \phi_i(a) > \sum_{i=1}^{n} \phi_i(b).$$

where for any city $x \in R$,

$$\phi_i(x) = \begin{cases} 1 & \text{if} \quad C_i(x) = + \\ 0 & \text{if} \quad C_i(x) = - \\ 0 & \text{if} \quad C_i(x) = ? \end{cases}$$

In the event that both numbers are the same, one simply guesses.

[25] In the weighted linear model, city $a$ is chosen over city $b$ if

$$\sum_{i=1}^{n} v_i \phi_i(a) > \sum_{i=1}^{n} v_i \phi_i(b),$$

where $v_i$ is ecological validity, and for any city $x \in R$,

$$\phi_i(x) = \begin{cases} 1 & \text{if} \quad C_i(x) = + \\ -1 & \text{if} \quad C_i(x) = - \\ 0 & \text{if} \quad C_i(x) = ? \end{cases}$$

In the event that both numbers are the same, one simply guesses.

are the best possible alternatives. This is an important drawback in the claimed superiority of fast and frugal heuristics over other possible solutions.

The main empirical results in Gigerenzer and Goldstein (1996) are as follows. Suppose that 100% of the cue values are known. In this case, the take-the-best algorithm generated as many correct inferences as any of the alternative algorithms under consideration. Now suppose that we consider the following percentages of known cue values: 10, 20, 50, 75, 100, and take the average number of correct inferences across all of them. Again, the take-the-best algorithm performs as well as any of the alternative algorithms used. This result is strengthened if we note that these figures ignore the costs of using the alternative algorithms. These costs include the relatively greater costs of acquiring and using additional information, and the greater computational time required relative to the fast and frugal take-the-best algorithm.

If only the recognition heuristic is used (i.e., choose the recognized object over an unrecognized object), then the accuracy across all pairs of choices is 65%, which is significantly above purely random choice. Finally, the authors find an intriguing *more is less effect*. Giving the decision maker more information (i.e., increasing the proportion of recognized objects) beyond some intermediate level reduces the accuracy of decision making.

Oppenheimer (2003) offers one criticism of the use of the recognition heuristic in the context of judging city size. He notes a problem with the original studies; they appeared to have conflated the size of the city with its recognition. For instance, Berlin is recognized and known to be large. The recognition heuristic predicts that in a comparison of the city size in which one city is recognized and known to be a small one, and the other city is unrecognized, the former should be chosen. By introducing names of fictitious cities so that many would not be recognized, Oppenheimer (2003) finds the opposite result: Namely, that the unrecognized cities are chosen over "recognized and known to be small cities". This suggests a more careful consideration of the robustness of the method and the claims.

The work of Gigerenzer and Goldstein (1996) is related to the work of Payne et al. (1993) who tested the performance of several heuristics in a choice between various risky gambles. They mainly looked at the following two one-decision heuristics that are very similar to the take-the-best heuristic. (1) The LEX heuristic in which cue order is determined not necessarily by ecological validity, unlike in the take-the-best heuristic, but by some measure of importance. (2) The LEXSEMI heuristic, similar to LEX except that the difference between cue values must exceed a certain threshold for a cue to be able to discriminate.

They found that different heuristics performed best in different environments, hence, the use of heuristics is strongly dependent on the environment, as Herbert Simon had proposed. Both LEX and LEXSEMI performed quite well in comparison with the weighted additive model (similar to the weighted linear model described above). Under severe time pressure, the LEX normally outperformed the others because it requires fewer computational steps.

The explanation of advertising has puzzled economists for a long time; none of the existing explanations in neoclassical economics is completely persuasive. However, the recognition heuristic might provide one important clue to an explanation. If consumers use the recognition heuristic in choosing among alternative products, then it is worthwhile for companies to spend significant sums of money on, often uninformative, but recognition-enhancing, advertising. The high-profile publicity campaigns of politicians, cities, or travel destinations can perhaps also be explained in a similar manner.

Gigerenzer and Goldstein (1996) take the cues as given; indeed these are often experimenter provided. In many contexts, the cues must be found in the first place. How do individuals search

for cues. One possibility is that individuals use fast and frugal heuristics to undertake the search in the first instance; see Todd (2001) for examples.[26]

In another intriguing study in the context of stock markets, Borges et al. (1999) pit the simple recognition heuristic against highly trained fund managers. Over the period 1996–1997, individuals were asked to invest in a variety of American and German stocks by using only the recognition heuristic. Namely, between two stocks, if one is recognized and the other is not, then choose the recognized stock; two of the four authors used their own money to fund this stock market experiment. In competition with the recognition heuristic were pitted several classical benchmarks for stock market selection such as mutual funds, market indices, and chance (or dartboard portfolios).

The authors found that the average level of stock market return on a randomly chosen portfolio was consistently below the return arising from the use of the recognition heuristic. Portfolios of highly recognized stocks outperformed those that were unrecognized. The performance of the recognition heuristic was also better (in six out of eight tests) relative to two major managed funds, the American Fidelity Blue Chip Growth Fund and the German Hypobank Investment Capital Fund. This is suggestive of a relationship between recognition, market share, and profitability of companies; Borges et al. (1999) also review some empirical evidence in this regard. These interesting results are only suggestive, and further empirical evidence is required. For instance, the results were obtained in a strong bull market and the time horizon for the performance of the stocks was only one year. These design features need to be relaxed in future work, certainly in a manner that most experimental economists will find persuasive.

Czerlinski et al. (1999) examine the performance of the take-the-best heuristic and the *minimalist heuristic* against statistical benchmarks such as the linear regression model. The minimalist heuristic is identical to the take-the-best heuristic except that at the first stage, the cues are not arranged in any particular order (such as ecological validity) but chosen randomly. They consider 20 different environments. One of these environments is the relative population sizes of two cities in a binary comparison (this is as in Gigerenzer and Goldstein, 1996). However, other environments include (all using US data): high school dropout rates, rate of homelessness, mortality rate, house prices, rents, and obesity.

In each of these environments, individuals are given sensible cues. For instance, when individuals were asked to predict the high school dropout rates in 57 Chicago public high schools, they were given the following cues: percentage of low-income students, percentage of nonwhite students, and average SAT scores. Take-the-best often outperformed multiple regression while employing only a third of the cues used by multiple regression. Even the minimalist heuristic came close to the accuracy of the multiple regression model. The relatively good performance of these simple heuristics seems to be the result of the excellent correlation of the cues with the target variable. However, the cues are experimenter determined and provided, which compromises the bite in the results.

Heuristics often appear to outperform multiple regression in a variety of contexts, and in particular, when one looks at out-of-sample prediction (Gigerenzer et al., 1999). The intuition is that multiple regression, while taking into account the salient features of the problem, also fits the noise in the sample. Consider an extreme case in which, outside the sample, such noise is absent or the nature of such noise is very different. Then there might be large prediction errors when an

[26] The example of a new colleague who joined our university from the US is instructive. He was interested in a broadband, TV, and telephone connection. He asked a few selected colleagues what they had chosen; the answers allowed him to immediately narrow down his search process quite substantially.

attempt is made to fit the multiple regression to out-of-sample points. On the other hand, simple heuristics do not respond to noise. By relying on the salient, they might be able to predict better outside the sample. However, this could also work to the detriment of the heuristics when there is no noise in the sample but much greater noise out of the sample.

In his "Mais lecture," Mervyn King (2005), the then Governor of the Bank of England, makes a case for using fast and frugal rules in central banking. His argument is that fully optimizing rules are too complex to fit in with the evidence on limited human cognition. Furthermore, as the state of the economy and the state of knowledge in economics changes, often at a rapid rate, problems of limited cognition and limited time become even more severe. Hence, an appropriate response is for both actors, the central bank and individual forecasters in the economy, to employ useful heuristics.

The central bank could adopt the following heuristic, depending on the environment. In normal times, set interest rates such that the expected inflation in two years time is equal to an inflation target. However, when the economy has been hit with a large shock, then the heuristic is to bring inflation back to target over a longer period. Hence, in this approach, the inflation target provides a nominal anchor around which the understanding of such a heuristic might become common knowledge. King also suggests heuristics that might be ideal (from the central bank's point of view) for the individual forecasters to use. The ideal heuristic is for individuals to expect inflation to equal its target. An example of a bad heuristic is that if inflation is different from the target, then it is expected to drift further away. These heuristics are fast and frugal, and expected to be relatively conducive to economic stability.

There is also a literature on using heuristics to make choices among objects with multiple attributes (Payne et al., 1993). Payne and Bettman (2001) consider the problem of choosing among several brands of cars with different attributes (e.g., reliability, price, safety, horsepower). A key part of the problem is that there is conflict among the various attributes, so none of the cars dominates the others on all attributes. Furthermore, decision makers could have several goals such as maximizing accuracy, ease of justification of the decision to others, minimizing cognitive effort, and negative decision-related emotions. Furthermore, decision makers seem to use different goals in different settings, and at different times. How should a decision maker choose?

Several strategies are considered in the heuristics literature. One possibility is the *weighted adding strategy* (WADD) in which the score of each car is added over all attributes using some subjective weights (Edwards and Tversky, 1967). Another heuristic that is computationally simpler is the *equal weighting strategy* (EQW) that assigns unit weights to all attributes. The LEX strategy (see above) simply chooses the most important attribute, say, reliability, and then chooses the car with the highest value on this attribute.

Simon's (1955) *satisficing strategy* (SAT) can also be used to make a choice in this case. Start with a base vector of attribute values that summarizes the aspiration levels for each attribute. Then pick the first car that does better than the base vector on all attributes. The *elimination by aspects* (EBA) strategy combines the LEX and SAT strategies. In this case, first choose the most important attribute and set an aspiration level for it. Then eliminate all cars that fall below this attribute value. Repeat this step for the next most important attribute, and so on, until a unique choice remains; if the choice is not unique, then some tie breaking rule is used (Tversky, 1972).

Payne and Bettman (2001) find that the heuristic which performs best is heavily contingent on the environment. The WADD strategy is the most robust in terms of accuracy across the various environments. But there are environments where the LEX achieves about 90% of the accuracy of the WADD with only about 40% of the effort. For more examples of fast and frugal heuristics and their performance, relative to statistical benchmarks, see Gigerenzer and Gaissmaier (2011).

We now offer a critical assessment of this approach.

1. The evidence provided by Gigerenzer and colleagues does not establish the conditions under which the heuristics proposed, such as take-the-best algorithm, will outperform the relevant statistical procedures or optimization approaches. They argue that the heuristics used by humans can outperform other statistical approaches to solving problems. This may often occur when the optimal benchmark is unclear. No heuristic can outperform an optimal solution when the optimal solution is unambiguously known. For instance, when the demand and cost curves are known, no heuristic can do better than the optimal solution: marginal revenues equal marginal costs. However, things are less clear if these curves are unknown or when one is interested in predicting stock market prices; in these cases the optimal solution with which to compare a heuristic is unclear.
   Furthermore, these results neither include the costs of finding the cues or the additional costs of information collecting, integrating, and computing associated with statistical procedures that are not fast and frugal; the latter issue is often considered in the theoretical computer science literature (Martignon, 2001).
2. In most studies of fast and frugal heuristics, the heuristics are proposed by the researchers or experimenters themselves. However, it needs to be established if individuals actually use these heuristics in actual practice. Some progress was made in this direction by Goldstein and Gigerenzer (1999). They gave individuals additional information that contradicted the decision that would have been made on the grounds of the recognition heuristic. They found that, nevertheless, 92% of the choices made by the participants still agreed with the recognition heuristic. It remains an open question whether individuals do follow the other fast and frugal heuristics considered above. In contrast to this, the thrust of the Kahneman–Tversky approach was to discover the "actual" heuristics that people follow.
3. How do individuals decide on the cues in a particular environment? One possibility is that the cues are searched using some simple heuristics. Perhaps we engage in simple learning that associates past cues with the outcomes. This is the basis of models of reinforcement learning that we consider in Part 5 (Erev and Roth, 2001). But it is also possible that we have, over the course of human evolution, acquired the ability to generate simple cues and heuristics. The problem of catching a ball (see Example 19.6) is a nice example of this phenomenon. However, it is less clear how these insights can be used to solve economic problems such as the design of contracts, choice of pension plans, choice of a mortgage, and choice of a monetary policy, for which evolutionary experience may provide only limited help. It is not immediately clear which fast and frugal heuristics to use in these cases. Unless these issues are directly addressed, they will limit the scope, and practical usefulness, of this work for economists.

## 19.15 The great rationality debate

One objection raised by some critics of Kahneman and Tversky's heuristics and biases program (henceforth, KT) is that it overly focuses on human biases and paints a pessimistic picture of human inference and prediction. By contrast, it is argued, the main thrust of the approach by Gerd Gigerenzer and colleagues (henceforth, G&C) is more positive because it shows that heuristics outperform more sophisticated forms of statistical reasoning. Kahneman and Tversky's approach came under great criticism only a few years after publication; see, Cohen (1981),

Einhorn and Hogarth (1981), Lopes (1991), Gigerenzer (1991, 1993, 1994), and Gigerenzer and Hoffrage (1995).

This debate between KT and G&C has come to be known in academic circles in psychology as the *Great Rationality Debate*; for surveys, see Stanovich and West (2000) and Stanovich (2012). Within psychology, the debate has been extremely acrimonious. Consider the following list of comments by psychologists on Kahneman and Tversky's work, taken from Stanovich and West (2000, p. 649):

Many critics have insisted that in fact it is Kahneman and Tversky, not their subjects, who have failed to grasp the logic of the problem (Margolis 1987, p. 158).

If a "fallacy" is involved, it is probably more attributable to the researchers than to the subjects (Messer and Griggs 1993, p. 195).

When ordinary people reject the answers given by normative theories, they may do so out of ignorance and lack of expertise, or they may be signaling the fact that the normative theory is inadequate (Lopes 1981, p. 344).

In the examples of alleged base rate fallacy considered by Kahneman and Tversky, they, and not their experimental subjects, commit the fallacies (Levi 1983, p. 502).

What Wason and his successors judged to be the wrong response is in fact correct (Wetherick 1993, p. 107).

Perhaps the only people who suffer any illusion in relation to cognitive illusions are cognitive psychologists (Ayton and Hardman (1997, p. 45).

These comments are sharp, some even vituperative. The discussion in Stanovich and West (2000) and the commentary following their excellent article illustrates just how muddled the debate really is. Part of the deadlock on these issues in psychology arises from disagreement over (1) the precise transmission channels in the brain through which judgment heuristics manifest (e.g., whether it is dual process theories or something else?), (2) the normatively correct model (as opposed to the rational benchmark in economics). These issues are not critical in economics so we omit them; interested readers can consult Stanovich and West (2000) and Stanovich (2012). Instead, we consider the remaining sources of disagreement below.

KT and G&C address very different questions. KT ask if people behave in the manner prescribed by the rational benchmark in economics. The rational benchmark implies compliance with subjective expected utility theory, and the classical rules of statistics and probability theory. The main finding in KT, that we extensively review in this chapter, is that a significant percentage of people do not act in a manner that is consistent with the rational benchmark.

G&C ask if there exist fast and frugal heuristics that still do better than "some statistical benchmarks" that rely on much greater information and computational requirements? They find the answer in the affirmative. However, typically little evidence is provided in this literature that people do indeed employ the experimenter-provided fast and frugal heuristic in real life. It is also not always clear what the proper statistical benchmark is against which a particular fast and frugal heuristic is compared.

Consider, for instance, the work that pits the recognition heuristic against several statistical benchmarks (see above) to determine which method earns greater stock market profits (Borges et al., 1999). Given that the efficient markets hypothesis is often rejected, the relevant benchmark that best predicts actual stock market prices is not clear. Indeed, economists and financial firms will not necessarily agree with the benchmarks employed in this literature. Hence, declaring victory for the recognition heuristic in this context may be premature. The only sensible inference that can be drawn is that for the set of benchmarks chosen by the experimenter and for

the chosen year and stock portfolio, the recognition heuristic did better. On the other hand, the full-optimization statistical benchmarks, even if they are known to exist, are likely to be computationally complex and informationally demanding. This is not a problem for financial firms who possess the computational firepower. However, for individual stock market participants with limited cognition and resources, the G&C approach poses appropriate questions, hence, imparting great usefulness to the approach.

Given the very different nature of the core issues dealt with in the KT and G&C frameworks, it is staggering that so many of the leading researchers could take such strong positions and fish in very muddy waters. Indeed, it is very likely that both KT and a less rhetorical G&C view of the world are, in fact, correct. The following two quotes from Kahneman (2000, p. 682) show just how close the two sides in the debate really are: (1) "Contrary to a common perception, researchers working in the heuristics and biases (HB) mode are less interested in demonstrating human irrationality than in understanding the psychology of intuitive judgment and choice." (2) "All heuristics make us smart, more often than not …".

There are, however, other, non-core issues that have to do with the nature of the evidence, where a debate is possible. A good source for this debate is Kahneman and Tversky (1996) and Gigerenzer (1996a). Some of the arguments against KT can be dismissed easily.

1. It has been argued that experimental evidence involving probabilities based on *singular events* is suspect because the frequentist approach requires a large number of repetitions (Gigerenzer et al., 1989). For instance, whether Linda is a bank teller or a feminist is a singular event. Whether a particular candidate will win the next Presidential elections in the US is also a single event. In this case, strictly speaking, one can speak only about subjective probabilities in a Bayesian sense (Gigerenzer et al., 1989). The argument is made that since only subjective probabilities are involved, any subjective probability will do, so the experimental results are uninformative about the violation of a rationality-based model. Kahneman and Tversky (1996) note that this issue only applies to 2 out of the 12 biases that they highlight. Furthermore, the responses show systematic biases in one direction, which weakens the case for arbitrary subjective probabilities.
2. It has been alleged that KT ignore the context, content, and framing effects in their demonstration of biases (Gigerenzer et al., 1988; Gigerenzer, 1993). However, this appears to be a misperception. Kahneman and Tversky (1996, p. 583) write: "the assumption that heuristics are independent of content, task and representation is alien to our position, as is the idea that different representations of a problem will be approached in the same way." Consider, for instance, the distinction between causal and non-causal base rates and their effect on the violations of base rate (see Section 19.7.1) that is explicitly recognized in Tversky and Kahneman (1980).
3. The argument has been made that subjects may simply be making errors, could be inattentive, or may be suffering from temporary lapses of judgment. Some of the suggestions that have been made to cast doubt on the KT program include temporary insanity of the subjects, difficult childhood of the subjects, and entrapment by the experimenter (Kahneman, 1981, p. 340). The pattern of biases discovered by KT does not describe random mistakes, but systematic mistakes, hence, these objections cannot be taken seriously. Furthermore, there is typically a great deal of heterogeneity among subjects who exhibit biases, which contradicts conformity with a single rational response. A more detailed discussion can be found in Stanovich and West (2000).
4. A popular criticism that Gilovich et al. (2002) describe as the "*we cannot be that dumb critique*" argues that stellar human achievements, e.g., discovering the structure of DNA

and space flights, are not consistent with the idea that we might be using simple judgement heuristics. This critique arises from a confusion about the nature and process of scientific discoveries. First, scientists might well be using simple heuristics to build initial intuition about their problems. Second, scientific progress is not judged against the standards of individual human judgment heuristics. Rather, it is delegated to an established scientific process that prescribes stringent empirical testing of theories. Hence, there is no essential contradiction between scientific progress and the use of judgment heuristics.

We now consider other criticisms of KT that require either a more detailed or subtle answer.

1. *Frequency versus probability format*: A prominent claim is that the biases identified by KT are eliminated to a large extent when data is presented in frequency format rather than probability format (Gigerenzer et al., 1988, Gigerenzer, 1991, 1994, 1996b). However, the distinction between the frequency and probability formats was first noted by Tversky and Kahneman (1983) themselves in the context of the conjunction fallacy. They found that the frequency format leads to very few conjunction errors; see Section 19.3. However, the conjunction bias is found to be sufficiently high in a between-subjects treatment, relative to a within-subjects treatment, even when data is presented in frequency format (Kahneman and Tversky, 1996).
   The anchoring heuristic was also tested by KT in a frequency format; see, e.g., the example on the number of African countries in the United Nations due to Tversky and Kahneman (1973, 1974) in Section 19.6. Indeed, strong evidence of anchoring was found despite the data being presented in terms of natural frequencies.
   In the context of base rate neglect, Table 19.4 in Section 19.7.1 shows that while the frequency format reduces base rate neglect, on average about 60% of the subjects in a cross section of studies nevertheless exhibit base rate neglect. Evans et al. (2000) report that the frequency format has been found, depending on the experiments, to worsen, improve, or leave unchanged, the quality of the judgments made.
   Thus, the weight of the evidence shows that while the frequency format improves compliance with the rules of classical statistics in some problems (e.g., in the conjunction rule), high non-compliance remains for most problems. Although humans might have evolved to cope better with natural frequencies, casual empiricism suggests that crucial real-world economic data is presented in percentage terms. For instance, interest rates on mortgage borrowing are quoted in percentage terms; the forecasts of inflation and unemployment by central banks are typically made in percentage terms; various insurable risks are often presented to potential insurees in percentage terms; grade information on open days in educational institutions is often given out in percentage terms; and league tables are typically based on inputs that are stated in percentage terms. Thus, it is important to study human judgment and perception in probability tasks.
2. *Formal definitions and statements of heuristics*: Another criticism of KT is that because the heuristics are not stated formally, it is not clear what constitutes a refutation (Gigerenzer, 1991, 1996a; Gigerenzer and Gaissmaier, 2011). For instance, talking of representativeness and availability, Gigerenzer (1991, p. 102) writes that these are: "largely undefined concepts and can be post hoc used to explain almost everything." Kahneman and Tversky's (1996) response is that since representativeness can be elicited experimentally (see Section 19.2.1) there is no need to define it a priori.
   However, an even better response is that we now have several models that formally define the exact bias in question and derive the formal implications. We have considered several

models in this growing class of models, above. These include: Rabin's (2002) model of inference by believers in the law of small numbers and Rabin and Vayanos's (2010) model of gambler's and hot-hands fallacies (see Section 19.2.3); the model by Biais and Weber (2009) on hindsight bias (see Section 19.8); and the model of Rabin and Schrag (1999) on confirmation bias (see Section 19.9). To this list must be added models of shrouded attributes and limited attention that consider explicit formalizations; see Sections 19.16, 19.17. We anticipate further progress in this direction.

3. *Narrow norms*: Gigerenzer (1996a) questions the appropriate statistical norms that apply to the problems in the experiments in KT; see also Koehler (1996). His argument is that the underlying model is not fully specified, which leaves the appropriate statistical prediction unclear. He writes (p. 592): "A convenient statistical principle, such as the conjunction rule or Bayes' rule, is chosen as normative, and some real-world content is filled in afterward, on the assumption that only structure matters. The content of the problem is not analyzed in building a normative model, nor are the specific assumptions people make about the situation."

   This statement should surprise most economists. The analogy in economics is the following. Either one can test the axioms of expected utility theory, e.g., the independence axiom, separately and without reference to an economic model; this often takes the form of observing choice in pairwise lotteries. Or, one can test the predictions of a model that has at its core, the model of expected utility. The first option is scientifically equivalent to the second option because the axioms of rationality are necessary and sufficient for an expected utility representation of preferences. However, the first option is the easier of the two because contrary evidence achieved in the second option is a joint rejection of expected utility and other auxiliary assumptions in the model. Indeed the first option has been massively employed in the testing of expected utility in economics. The airframe of a proposed new airplane might be tested in a wind tunnel, without adding-on all the other components in the airplane. This is a scientific, acceptable and sensible procedure.

   Thus, Gigerenzer, in this case, criticizes an appealing feature of the KT tests, i.e., clean unambiguous predictions. Most reasonable economists would be comfortable with the predictions ascribed to the relevant statistical models in the KT tests; although they might contest issues of experimental design such as the appropriateness of subject pools and the appropriate incentivization of the subjects.

The conclusion reached by Kahneman and Tversky (1996, p. 584) is blunt: "The position described by Gigerenzer is indeed easy to refute but it bears little resemblance to ours. It is useful to remember that the refutation of a caricature can be no more than a caricature of a refutation." However, on another count, G&C need to be commended for asking fundamentally important questions about the usefulness of heuristics and making important headway in this direction. KT are likely to agree with this position, although they would be dismayed at how muddled and contentious the debate between the two sides has become.

Both sides also advocate and vigorously debate models of the brain that may lead to biases relative to statistical benchmarks. Kahneman (2011) devotes the first one-third of his book to developing such models of the brain, in this case a model of Systems 1 and 2, that facilitates a deeper understanding of the biases.[27] For instance, the quick, reactive, and automatic System 1 is responsible for most errors. System 2 has been likened to a *lazy controller* in Kahneman (2011)

[27] However, there is no universal agreement among psychologists about the appropriate models of the brain in this context (Stanovich and West, 2000).

and when it is called upon to intervene in an unusual situation, the agenda (e.g., affective emotions, recalled memory, associations) is chosen by System 1. On p. 21 he writes: "Although System 2 believes itself to be where the action is, the automatic System 1 is the hero of the book." System 1 tries to make sense of a situation even when the events may have been generated purely randomly. As Stanovich (1999) points out, System 1 has a tendency toward automatic contextualization of problems. This gives rise, for instance, to the law of small numbers, the gambler's and hot hands fallacies, and overconfidence in the ability to explain the future from a hindsight-biased explanation of past events that were purely random.

There should be no presumption that evolution must have favored rationality, as defined in neoclassical economics. Natural selection chooses among a population phenotype and a mutant phenotype, on the basis of which of the two adapts better to the environment. Indeed traits like loss aversion, prosociality, hyperbolic discounting, and level-$k$ reasoning might be favored in this selection process. There is no presumption that evolutionary mechanisms impart greater rationality to individuals, understood as subjective expected utility maximization, or conformity with the rules of probability theory (Stanovich, 2012).

## 19.16 Shrouded attributes

Attributes are *shrouded* when some characteristics of the goods that consumers buy are not directly observable. These hidden attributes can typically be discovered by the consumers at some cost, which can be a monetary cost or a purely cognitive search cost. For instance, while opening a bank account, an individual might not know the minimum balance fees or bank surcharges on using its cash machines, or bounced check fees. Neither does the bank make an attempt to easily provide this information, even on its website, although it is tucked somewhere in the fine print, or in some unexpected corner of the website. Manufacturers of printers shroud information on the costs of their patented ink cartridges that can, over the printer's lifetime, cost several times as much as the cost of the printer.

In classical theory, the costs of *unshrouding* the shrouded attributes are either considered so small that all attributes are assumed to be visible to the consumer, or competition among firms leads to the unshrouding of attributes. Consumers who are rational will believe that shrouded attributes signal bad news and attempt to find other sellers. In response, firms will have an incentive to unshroud information (Jovanovic, 1982; Milgrom, 1981).

### 19.16.1 *Empirical evidence*

In actual practice, one observes that individual decisions are often visibly influenced by shrouded attributes, even when the cost of unshrouding them appears to be quite small. Consider, for instance, two state taxes on alcohol in the US. The *excise tax* is included in the price of alcohol, while the *sales tax* is not included in the tagged price but it is added at the point of purchase. In classical theory, both taxes should have identical effects on the demand for alcohol. However, the sales tax is shrouded, while the excise tax is unshrouded. Figure 19.11 shows that the elasticity of demand of alcohol with respect to the two taxes is very different. Indeed, as the reader might expect, demand is relatively less responsive to changes in the shrouded tax, relative to the more visible tax.

Consider another example. Hossain and Morgan (2006) conduct a field experiment on eBay auctions in the US in which they put CDs and XBox games up for sale for different levels of

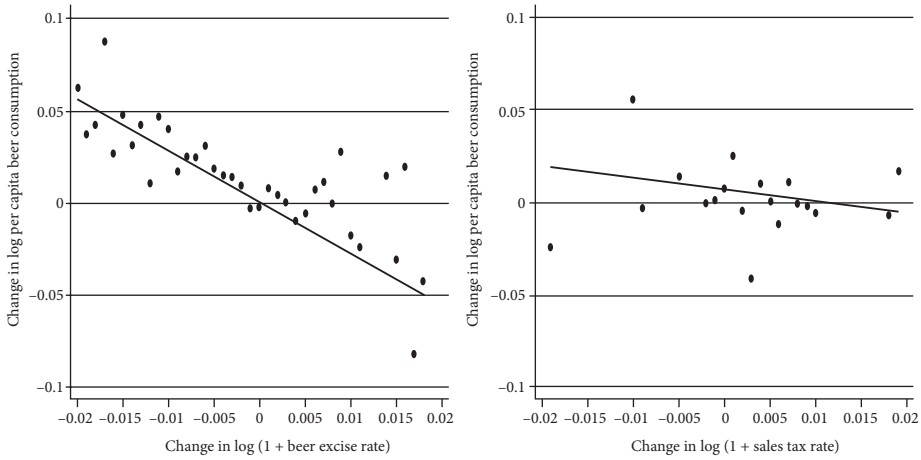


**Figure 19.11** Variations in demand due to a visible excise tax (left panel) and a shrouded sales tax (right panel).

Source: Chetty et al. (2009) with permission from the American Economic Association.

shipping charges and initial bids. The expectation is that shipping charges are shrouded attributes. eBay auctions are an example of ascending bid English auctions and standard auction theory predicts an efficient outcome. Suppose that a bidder has a maximum willingness to pay of $\$x$ for an item. If shipping and handling were free, the bidder should be willing to pay up to $\$x$. Suppose that the minimum opening bid is $\$m$, and shipping and handling costs are $\$s$, and $x \geq m + s$, where $m + s$ is known as the *effective reserve price*. Then, the bidder would be willing to bid $\$b \geq \$m$ for the object, up to $b + s = x$. In particular, so long as these inequalities hold, the buyer should be willing to part with a total of $\$x$ for the object, even as $m$ and $s$ vary.

Two treatments were run. In one treatment, the effective reserve price, $m + s$, was low, and in another it was high, but in each case, $x \geq m + s$. In each treatment, different levels of $m$ and $s$, keeping fixed $m + s$, were tried out. The predicted revenues of the seller are identical. In contrast to the predicted results, the main empirical finding is that in both treatments, low $m$ and high $s$ led to more bidders, earlier bidding, and higher revenues for both CDs and Xbox games. Thus, it appears that an increase in shipping and handling charges, keeping the effective reserve price constant, increases bidding and seller revenues. This is consistent with the view that many buyers treat shipping and handling charges as shrouded attributes. These results are not unique to eBay auctions in the US and have been replicated in leading online auction platforms in Taiwan and Ireland (Brown et al., 2010).

## 19.16.2 *A theoretical framework*

Gabaix and Laibson (2006) develop a general framework to model market competition when some attributes of a good may be shrouded. We have noted above that in the neoclassical case with rational consumers, no shrouding is predicted to take place in equilibrium. However, when not all consumers are rational in the neoclassical sense, we show that shrouding of attributes can take place in equilibrium. Furthermore, firms do not have an incentive, in equilibrium, to educate consumers about the shrouding practices of their rivals.

Consider the following example that nicely illustrates these claims. Suppose that there are two components of the price of a hotel room. The room price or the base price, $p_b$, and the price of add-ons, $p_a$, such as costs of parking, telephones, internet connections, and room service. The total price paid by the consumer is $p = p_b + p_a$. The hotel does not keep $p_a$ secret from the public, but makes it sufficiently inaccessible that one needs to expend non-trivial search costs to find it.

The hotel sector is assumed to be perfectly competitive. Suppose that a fraction $\alpha \in [0,1]$ of the customers is *myopic* (myopes) and the remaining fraction $1 - \alpha$ is *sophisticated*. Myopic consumers cannot observe shrouded information.[28] However, myopes exhibit heterogeneity in the presence of unshrouded information. Under unshrouded information, a fraction $\lambda \in (0,1]$ of the myopes observes the information (informed myopes) and the remaining fraction $1 - \lambda$ (uninformed myopes) pays no attention to unshrouded information about add-ons. Sophisticates comprise a fraction $1 - \alpha$ of the consumers and take account of add-ons when they are unshrouded.

Suppose that one of the hotels is the Hilton and it costs \$100 for the hotel to supply the room. Assume that add-ons have zero costs to the Hilton but consumers typically spend \$20 worth of add-ons when they have to pay for them. In competitive markets, the Hilton will set $p_b = 80$ and $p_a = 20$ and make no profits. Thus, add-ons have the advantage that they subsidize the base price. Consider a competitor hotel, called the Transparent, that supplies free add-ons and also advertises/educates consumers about the add-on prices at its competitors. We assume throughout that education of consumers is costless to the firm. Since the Transparent operates in competitive markets, it must set $p_b = 100$ and $p_a = 0$, thus, the add-ons price cannot subsidize the base price. Let us consider the behavior of each kind of consumer.

1. Uninformed myopes: These consumers find the Hilton base price cheaper. Since they ignore the price of add-ons, they choose the Hilton.
2. Informed myopes and Sophisticates: These consumers benefit from the information that is provided by the Transparent. So they realize that they will spend an amount \$20 on add-ons at the Hilton. They might be able to arrange for these add-ons from the outside. For instance, take a taxi to the hotel to avoid paying parking and take a cell phone to avoid hotel call charges. Suppose that this substitution effort costs them $e = \$10$. These consumers will find it profitable to book a room at the Hilton and pay the base price of \$80 plus the substitution effort cost of \$10 for a total of \$90. This works out cheaper than the price of \$100 for booking a room in the Transparent hotel.

Thus, in equilibrium, the Transparent does not benefit from its policy of advertising about add-ons. By contrast, if the Transparent shrouds add-ons, uninformed and informed myopes may book rooms with it. Unlike the classical framework with purely rational consumers, the presence of myopic consumers allows some attributes to be shrouded in equilibrium.

Consider a formal version of the two-period problem given above in which competitive hotels are indexed, $i = 1, 2, \ldots$

In period 1, hotel $i$ offers a base good at a price $p_b^i$ and the price of add-ons is denoted by $p_a^i$. Myopes comprise a fraction $\alpha$ of consumers, and a fraction $\lambda$ of myopes are *informed myopes* who take account of add-ons if unshrouded. Unshrouding is costless to the firm and takes the

[28] One possibility is that myopic consumers simply do not pay enough attention. Empirical evidence is suggestive of the importance of consumers who exhibit limited attention; see, for instance, DellaVigna and Pollet (2009), Cohen and Frazzini (2008), Hirshleifer et al. (2009), Gilbert et al. (2012).

form of educating consumers about the add-ons. Sophisticates who comprise a fraction $1-\alpha$ take account of add-ons when unshrouded.

Informed myopes and sophisticates form Bayesian posteriors, $Ep_a^i$, of add-ons. When add-ons are unshrouded, informed myopes and sophisticates can take account of add-ons, either by (i) choosing a firm that offers cheaper add-ons, or (ii) substitute away from the add-ons at a substitution cost $e : 0 < e < \overline{p}$; $\overline{p}$ can be thought of as the maximum price for add-ons that is allowed by government regulation. Informed myopes, like sophisticates, exert substitution effort only if $e < Ep_a$.

We only consider a symmetric price equilibrium. Denote by $v_i$, the surplus that a consumer gets by booking at hotel $i$, net of the opportunity cost of booking at the next best hotel. Variables corresponding to the next best alternative are starred. Thus, for informed myopes and sophisticates, the surplus is[29]

$$v_i = \left(-p_b^i - \min\left\{e, Ep_a^i\right\}\right) - \left(-p_b^* - \min\left\{e, Ep_a^*\right\}\right). \tag{19.77}$$

By contrast, uninformed myopes do not take account of add-ons, so they never exert any substitution effort. Thus, their surplus is given by

$$v_i = \left(-p_b^i\right) - \left(-p_b^*\right) = p_b^* - p_b^i. \tag{19.78}$$

Let $x(v_i) \in [0,1]$ denote the probability that a consumer books a room in hotel $i$; it captures the demand by the consumer. We assume that $x'(v_i) > 0$, so if a hotel offers better value it is more likely to be chosen.

In period 2, consumers who have booked hotel $i$ can purchase add-ons at a price $p_a^i$. If consumers have engaged in substitution efforts, then they do not purchase add-ons. We now state the main result.

**Proposition 19.5** *Consider a symmetric equilibrium in which the belief of sophisticates and the informed myopes is that if hotel i shrouds add-ons, then* $p_a^i = \overline{p}$.

*(i) Shrouded prices equilibrium: If* $\alpha > \frac{e}{\overline{p}}$, *then the price of the base good and the price of add-ons are given, respectively, by*

$$p_b^i = -\alpha\overline{p} + \frac{x(0)}{x'(0)} \text{ and } p_a^i = \overline{p};\ i = 1, 2, \ldots \tag{19.79}$$

*In this equilibrium, only the myopes purchase add-ons.*

*(ii) Unshrouded prices equilibrium: If* $\alpha < \frac{e}{\overline{p}}$, *then the price of the base good and price of add-ons are given, respectively, by*

$$p_b^i = -e + \frac{x(0)}{x'(0)} \text{ and } p_a^i = e;\ i = 1, 2, \ldots \tag{19.80}$$

*All consumers purchase the add-ons in this equilibrium.*

Proof: We sketch a part of the proof for (i) and leave the rest as an exercise. We are interested in computing a sequential equilibrium of the game. Thus, we need to check for (a) sequential

[29] We can, without loss of generality, include a benefit $B$ of staying in a hotel, but it cancels out in the subsequent calculation (unless there is a hotel-specific benefit of booking a room).

rationality, and (b) consistency of beliefs and actions. If a firm chooses to unshroud (i.e., educate the consumers about the add-ons, so that the informed myopes behave like sophisticates), then the beliefs of the sophisticates and informed myopes are the announced value of $p_a^i$. Otherwise, if the firm chooses to shroud, then given the stated beliefs, sophisticates believe that $p_a^i = \overline{p}$, i.e., once a consumer is locked-in with a hotel it will be charged the maximum permissible price for the add-ons.

(i) Suppose that $\alpha > \frac{e}{\overline{p}}$ and all firms shroud. We need to check that no profitable deviations from the announced strategies are possible. Given the prices of other firms, $p_b^*$ and $p_a^*$, if firm $i$ shrouds, then the myopes do not show any awareness of add-ons, and given the stated beliefs, consumers believe that $p_a^i = \overline{p}$ for all firms. Then, from (19.77), (19.78), we get $v_i = p_b^* - p_b^i$ for sophisticates and myopes. Thus, the demand for firm $i$ is given by $x(p_b^* - p_b^i)$ and the fraction $\alpha$ of myopes buy the add-ons. The profits of firm $i$, therefore, are given by

$$\pi = (p_b^i + \alpha p_a^i)x(p_b^* - p_b^i). \tag{19.81}$$

Firm $i$ chooses $p_b^i$ and $p_a^i$ to maximize $\pi$, given in (19.81). The solution to $p_a^i$ is directly obtained as the maximum possible value $\overline{p}$. Setting $p_a^i = \overline{p}$ in (19.81) we get $\pi = (p_b^i + \alpha\overline{p})x(p_b^* - p_b^i)$. Hence, the first order condition with respect to $p_b^i$ is

$$-(p_b^i + \alpha\overline{p})x'(p_b^* - p_b^i) + x(p_b^* - p_b^i) = 0. \tag{19.82}$$

In a symmetric equilibrium, $p_b^i = p_b^*$. Substituting this in (19.82) we get $p_b^i = -\alpha\overline{p} + \frac{x(0)}{x'(0)}$. This is the solution given in (19.79). However, to prove that this is the optimal solution to the problem, a bit more work is needed. We need to consider the case where the firm chooses to unshroud (i.e., educate consumers of the add-ons). Unshrouding educates the sophisticates and the informed myopes who might or might not respond by engaging in a substitution effort, $e$, depending on whether the price of add-ons is higher or lower. Thus, in this case we need to consider two cases: $p_a^i > e$ and $p_a^i < e$. In particular, the fraction of myopes in this case is given by $\alpha_1 = (1 - \lambda)\,\alpha$ because the informed myopes (fraction $\lambda$ of the total myopes) now behave like sophisticates. We leave it to the reader to consider these two cases and show that in each case, the profits of the firm are lower relative to the case of shrouding.

(ii) Proof left to the reader as an exercise. ■

Proposition 19.5 illustrates two different equilibria. In Proposition 19.5(i), firms choose to shroud and charge the maximum permissible price for add-ons, $p_a^i = \overline{p}$ and subsidize the base price relative to the unshrouded equilibrium in Proposition 19.5(ii). Empirical evidence for this prediction is reviewed in Gabaix and Laibson (2006) and Ellison (2005). The intuition for charging the maximum permissible price for add-ons is that shrouding induces the worst possible beliefs about the price of add-ons. Since $e < \overline{p}$, sophisticates engage in substitution effort. Thus, only the myopes buy add-ons, and once they book into the hotel, the hotel might as well charge them the maximum allowed for add-ons.

This is the sense in which the equilibrium in Proposition 19.5(i) is not efficient because only the myopes purchase the add-ons. Furthermore, the sophisticates take advantage of the subsidized base price to improve their surplus under shrouding relative to no shrouding.

**Proposition 19.6** *The consumer surplus of the sophisticates is higher under shrouding relative to unshrouding when there is perfect competition among firms.*

Proof: Since perfectly competitive firms face an infinitely elastic demand curve, so $\frac{x(0)}{x'(0)} = 0$. From (19.79), (19.80) the optimal base price is $p_b^i = -\alpha\overline{p}$ under shrouding and $p_b^i = -e$ under unshrouding. Thus, the surplus of sophisticates under shrouding is $-p-e = \alpha\overline{p}-e$; since $e < \alpha\overline{p}$, the surplus is strictly positive. However, under unshrouding, the surplus is $-p-e = e-e = 0$. ■

The fact that sophisticated consumers prefer to engage with firms who shroud and charge high add-on prices but subsidize by offering low base prices is known as the *curse of debiasing*. An example is the low price of printers and the relatively high prices of patented printer cartridges. Sophisticated consumers, in this case, can buy unbranded low-priced printer cartridges on the Internet. In the neoclassical case with no myopes, advertising by firms will eliminate shrouding (Shapiro, 1995). However, the presence of myopes ensures that a captive set of consumers will pay a high price for add-ons in the presence of shrouding. We can see from Proposition 19.5 that if there are no myopes ($\alpha = 0$), then the condition $\alpha > \frac{e}{\overline{p}}$ required for the shrouded equilibrium in Proposition 19.5(i) is never satisfied.

By contrast, in Proposition 19.5(ii), the firm chooses to unshroud and chooses a price for add-ons that just makes consumers indifferent between substitution or not. We assume that the tie breaking rule is that in this case they purchase the add-ons (otherwise the price can be dropped by a penny to ensure that this occurs). This case arises when $\alpha < \frac{e}{\overline{p}}$, so the fraction of myopes is too low to make it worthwhile to extract sufficient consumer surplus by high prices of add-ons; this makes shrouding ineffective.

## 19.17 Limited attention

In this section, we consider the empirical evidence on *limited attention*. We also outline a general theoretical framework to model this phenomenon.

### 19.17.1 *Limited attention and the poor*

Why do individuals take payday loans with associated annual interest rates as high as 400%? Bertrand and Morse (2011) locate the reason in the limited attention of borrowers. They run several treatments in which potential borrowers are provided with a range of information on the real costs of payday loans. The information includes the weekly accumulation of fees (e.g., $270 in 3 months on a $300 loan), comparison of borrowing costs from other financial institutions, and refinancing using their credit card borrowing. These interventions succeeded in reducing the extent of borrowing; in the four treatments, the percentage reduction in borrowing was 5.9%, 12%, 16%, and 23%, respectively. These results suggest a potential role for legislation in making the provision of such information mandatory. However, the relatively small reductions in borrowing also indicates that limited attention alone does not account for the take-up of payday loans.

Stango and Zinman (2015) find that drawing attention to fees for overdrawn accounts, reduces overdraft fees for up to 2 years since the first reminders are sent out. The effects are more pronounced for individuals who have lower education and financial literacy. Limited attention might be expected to play a greater role when the number of options is greater. In an empirical study of the take-up of loans, Bertrand et al. (2010) find that reducing the number of options from four to one, increased loan take-ups; the effect is equivalent to a 2.3% reduction in the interest rate.

In Mexico City, only 39% of low-income individuals could spot the lowest cost credit product in bank brochures. However, 68% could spot the lowest cost product in a user-friendly summary sheet (Giné et al., 2014); limited attention and shrouded attributes are possible explanations. These examples also highlight the need for financial market regulation for framing information in a more transparent manner in the presence of limited information.

Limited attention is likely to reflect the cognitive costs of engaging in greater attention. These cognitive costs, which might take the form of depletion of scarce willpower required to exert greater attention, may create undesirable outcomes, especially for the poor. The poor may need to expend a disproportionate amount of scarce attention and willpower costs on important daily issues. These issues include arranging food for the table, dealing with debt, or arranging for clean drinking water; the associated willpower costs may be termed as a form of *cognitive tax* (Shah et al., 2012; Mullainathan and Shafir, 2013).

The result of this depletion in willpower is that the poor are more likely to engage their System 1 in making other potentially important decisions; the control exerted by System 2 on System 1 falls as willpower depletes. Thus, the poor may engage in insufficient deliberation on many important issues (Banerjee and Mullainathan, 2008); this is sometimes referred to as the problem of *low bandwidth*. Hence, poor individuals may make bad decisions, which perpetuate poverty that is the cause of the initial problem.

Indeed, the actions of the poor may not fully reflect their intentions; this is known as the *intention-action divide*. In the case of the poor, this problem is mostly caused by (1) intense willpower depleting focus on pressing current problems, and (2) present-biased preferences. Mani et al. (2013) find that farmers score much lower on cognitive ability tests before a harvest when they have high levels of financial stress. Measured IQ is 10 points lower before the harvest, relative to the period after the harvest.

Karlan et al. (2012) find that text messages sent out to individuals in Peru, Bolivia, and the Philippines, reminding them of their savings goals, actually improve savings by 16%. Presumably, these messages work because of issues of limited attention and procrastination (see Part 3 of the book). Furthermore, the most effective messages were the ones that emphasized a particular goal, for instance, purchase of a house or an appliance. Limited attention might also explain why many sophisticated financial literacy programs are not successful. Drexler et al. (2014) found that reducing the degree of sophistication of the financial programs, by emphasizing simple heuristics, improves financial literacy of firms and micro-entrepreneurs in the Dominican Republic.

The solutions to the problems of cognitive tax and low bandwidth can sometimes be surprisingly simple and offer a low-cost means of bringing about large welfare changes for the poor; Datta and Mullainathan (2014) term this as *low hanging policy fruit*. This suggests that poverty is not simply a problem of insufficient resources, so this offers a radically new perspective on the problem (WDR, 2015). Development policy could then play a key role in the timing of important decisions. For instance, sugarcane farmers have lower consumption in the months prior to the harvest and they may even pawn jewelry. However, if part of their harvest money could be put into an account with pre-specified withdrawal rates, then they may better smooth their consumption levels. Such default options have been shown to have important effects in the West for pensions and savings.

In the presence of limited attention and limited cognitive abilities, a benevolent or cunning third party might wish to frame the context within which individuals make decisions. This is known as the problem of *choice architecture*. Framing of context can take many forms; for instance, which information to make salient, which comparisons among options to highlight, which anchors to choose and so on. One example of choice architecture is Thaler and Sunstein's

(2008) idea of a *nudge*. For instance, on account of limited attention, if the healthiest snacks are put at eye level, they are more likely to be chosen; these issues are considered in more detail in Part 8.

## 19.17.2 *Limited attention and taxes*

Does consumer demand respond more to easily observed taxes relative to less salient taxes? For instance, sales taxes may be (i) posted on price stickers in shops, or (ii) not posted on the price sticker but added at the sales register at the point of sale. If demand or its responsiveness turns out to be different in the two cases (see Figure 19.11), then two competing explanations arise.

First, consumers may be unaware of the sales tax. Second, hidden fiscal attributes are less salient. Chetty et al. (2009) surveyed consumers at a grocery store where the sales tax was displayed for some but not other consumer goods. They found that the median consumer was able to successfully identify the sales tax on seven out of eight products on the survey. This finding, and the fact that consumers could easily inquire about the level of sales tax from the shop attendant if they wished, suggests that taxes that are not displayed on price stickers reduce tax salience.

Consider a simple model of consumer demand. Suppose that there are two consumer goods, $x$ and $y$. Good $y$ is an untaxed good whose price is normalized to 1. The pretax price of good $x$ is $p$ and the post-tax price is $q = (1 + \tau)p$, where $\tau$ is an ad-valorem sales tax. Consumers observe the pre-tax price on the posted price stickers in shops, but the sales tax, $\tau p$, is not included in the posted price.

Denote the solution to the consumer's optimal demand for $x$ by $x(p, \tau)$. In traditional public finance theory, the consumer takes account of the post tax price, $q$, even if the sales tax is not posted on the price sticker. Thus, we have $x(p, \tau) = x((1 + \tau)p, 0)$, so there is no need to separately observe the sales tax if the consumer already takes account of the post tax price. In this formulation, an increase in the post tax price, $q$, whether it is caused by an increase in $\tau$ or an increase in $p$, has identical effects on the demand for $x$.

Let $\epsilon_{x,p}$ and $\epsilon_{x,1+\tau}$ denote, respectively, the elasticity of demand for $x$ with respect to $p$ and $1 + \tau$. Under classical public finance theory we should, by definition, observe that

$$\epsilon_{x,p} \equiv -\frac{\partial \log x}{\partial \log p} = \epsilon_{x,1+\tau} \equiv -\frac{\partial \log x}{\partial \log(1+\tau)}. \tag{19.83}$$

In contrast to (19.83), Figure 19.11 shows that $\epsilon_{x,p} > \epsilon_{x,1+\tau}$. One way of representing limited attention to taxes that are not displayed on price stickers is to write $x(p, \tau)$ in a log-linear form.

$$\log x(p, \tau) = \alpha + \beta \log p + \theta\beta \log(1 + \tau). \tag{19.84}$$

In (19.84), the coefficients $\beta$ and $\theta\beta$, respectively, give the two elasticities defined in (19.83). Thus,

$$\theta = \frac{\frac{\partial \log x}{\partial \log(1+\tau)}}{\frac{\partial \log x}{\partial \log p}} = \frac{\epsilon_{x,1+\tau}}{\epsilon_{x,p}}. \tag{19.85}$$

If $\theta = 1$, then we have the classical case in (19.83). However if $\theta < 1$, then we get the observed result in Figure 19.11. If one could uncover independent variation in $p$ and $\tau$, then $\epsilon_{x,p}$ and $\epsilon_{x,1+\tau}$ can be estimated separately to compute $\theta$.

Chetty et al. (2009) conducted an experiment in a supermarket for three weeks in 2006. Initially, all products in the store excluded the sales tax on the price sticker; the sales tax was added later at the point of purchase. The price tags on a selected list of 750 products were altered to separately display the pre-tax price and the sales tax of 7.375% in order to increase the salience of the sales tax. The products included three groups: cosmetics, deodorants, and hair care accessories. The posting of the tax on the price stickers reduced demand by 8% relative to two control groups: other products in the store and similar products in two nearby stores.

The reduction in demand was statistically significant. One can reject the null hypothesis that posting of the sales tax on price stickers has no effect by using a t-test and non-parametric permutation tests. The effect of tagging products with the sales tax caused a fall in demand that was identical to a 7.373% increase in price. Hence, consumers pay limited attention to taxes that are not displayed on price stickers. $\theta$ is estimated to be 0.35. A separate survey conducted on supermarket shoppers showed that the median shopper was aware of the level of sales tax. Hence, the observed effect on demand was not the result of dispelling of ignorance, but an increase in the salience of the sales tax.

In a second study, the authors consider separately the effect of different kinds of indirect taxes on alcohol consumption. They considered the effect of excise duties that are included in the alcohol price and sales taxes that are only included at the point of sale, hence, these are shrouded; see Figure 19.11. The main empirical result is that, based on data over the period 1970–2003, excise taxes (more visible) reduced alcohol consumption far more than sales taxes (shrouded).

Goldin and Homonoff (2013) find that attention to sales taxes levied at the sales register differs among low-income consumers (who pay greater attention) relative to high-income consumers. This can be exploited by the fiscal authorities to reduce the degree of perceived tax regressivity by taxing goods consumed by the rich, relatively more.

Results of a similar nature have been obtained elsewhere. For instance, the earned income tax credit (EITC) is the largest cash transfer program in the US that transfers income to low-income families. However, evidence suggests that the marginal incentives implied by the program, e.g., should one work an extra hour, are not salient, hence, they would be ignored by individuals with limited attention. The EITC program appears to have had a relatively small impact on the intensive margin, i.e., hours worked, relative to the extensive margin, i.e., the decision to participate or not in the labor force (Hotz and Scholz, 2003).

Chetty and Saez (2013) arrange for tax experts to provide information on the intensive margin for potential EITC beneficiaries. Information provision takes several forms: a verbal description; a graph that reveals the behavior of the EITC as earnings change; and a table that lists key information on the EITC . Furthermore, potential beneficiaries were given advice that was particularly tailored to their personal circumstances. For instance, they were told if it would pay for them to work more or not. Yet, despite these attempts, the average behavior of the potential beneficiaries along the intensive margin did not change. One potential reason why providing information does not work in this case is that this problem is relatively more cognitively challenging relative to simply adding a sales tax on the price tag. Another reason is that the intensive margin may be determined by factors entirely outside this model.

Buchanan (1967) and Brennan and Buchanan (1980) note the potential link between tax salience and the size of the government. The basic idea is that governments can expand by increasing taxes that are not salient, ruling out adverse electoral consequences for political parties.

Finkelstein (2009) confirms these findings by examining the effect of electronic toll collections (ETC) in car plazas that automatically deduct the relevant toll; there is no manual exchange of cash between two parties in this case. Comparing customers who pay by manual tolls and by ETC, she finds much less awareness of the level of the toll among the ETC customers.

She also finds that there is an increase in the toll rates (the analog of an increase in the size of the government) following the introduction of an ETC. Indeed, toll rates are 20–40% higher in the presence of ETC as compared to manual toll collections. Furthermore, the timing of ETC increases is less sensitive to the electoral calendar relative to manual tolls, suggesting that politicians are less concerned about its electoral implications, most likely because people pay limited attention to these changes.

### 19.17.3 *Biased numerical attention*

It has been noted that people might unduly focus on the leftmost digit of a multi-digit number, relative to the digits on the right (Korvost and Damian, 2008). This appears to be a form of imperfect attention to some digits over others; we may term it as the *left-digit bias*. Basu (1997, 2006) explained the prevalence of prices ending in the digit 9; for instance, a price of $\$x.99$ (where $x$ is a whole number), rather than an integer price of $\$x + 1$; e.g., a price of \$39.99 rather than \$40. He shows that in a rational expectations equilibrium, all firms choose prices ending in 99 cents and consumers, rationally, do not benefit by observing the full price (say, \$3.00, when the quoted price is \$2.99). This explanation is quite distinct from the one that ascribes such biases to limited attention arising from cognitive limitations.

Lacetera et al. (2012) explore the presence of the left-digit bias in the wholesale used car market. Potential buyers may exhibit the left-digit bias when presented with information on the odometer readings of cars. In particular, they find that there are discontinuous drops in the retail prices of cars at 10,000 mile thresholds. For instance, cars with odometer readings 79,900–79,999 miles sell for (i) \$210 more than those with odometer readings in the range, 80,000–80,100, yet only (ii) \$10 less than those with odometer readings 79,800–79,899. There are also discontinuities at the 1000 mile thresholds but these are smaller and we shall ignore them here. Furthermore, there are spikes in the volume of cars presented for sale just before they hit these odometer thresholds, suggesting awareness of this bias among car sellers. These results are also shown to hold for the retail used car market (Busse et al., 2013).

Buying a car is a large stakes decision and odometer readings are easily observed. Hence, this evidence is an important demonstration of the presence of heuristics in real markets. As with all heuristics, individuals commit systematic biases relative to the benchmark of classical statistics. The left-digit bias may be a useful heuristic in the fast-and-frugal sense, and this needs to be explored in future work. For instance, buying a car just below the threshold is costly because it will soon cross the threshold but the benefit might arise if this information is generally correlated with other useful characteristics of the car.

Consider a formal model of left-digit bias that illustrates the limited attention aspect of the bias. Suppose that the reading on the odometer is a $k$ digit number, denoted by $m = d_1 d_2 \ldots d_k$, where $d_1$ is the leftmost digit and $d_k$, the rightmost digit; each digit belongs to set $\{0, 1, \ldots, 9\}$. Suppose that $m$, the actual odometer reading, is a positive whole number less than or equal to 100,000. A potential buyer perceives the odometer reading as $\hat{m}$, where[30]

[30] A slightly different formula applies when $m > 100{,}000$ because the changes in this case are not to the leftmost digit until $m$ exceeds 200,000. The reader may wish to think about the necessary modifications in the formula.

$$\hat{m} = d_1 10^{k-1} + (1-\theta)\sum_{j=2}^{k} d_j 10^{k-j}, \tag{19.86}$$

where $\theta \in [0,1]$ is the *inattention parameter*; thus the weight put on the leftmost digit, 1, is greater than the weight that is put on any of the other digits, $1-\theta < 1$. The classical case corresponds to full attention, $\theta = 0$. An individual is maximally inattentive if $\theta = 1$, in which case, only the leftmost digit is salient. For instance, for $m = 4326$, (19.86) implies that $\hat{m} = 4 \times 10^3 + (1-\theta)\left(3 \times 10^2 + 2 \times 10^1 + 6 \times 10^0\right)$. Whenever $0 < \theta \leq 1$ we have $\hat{m} < m$, i.e., the perceived mileage is lower than the actual mileage.

The actual value of the car is $V = V(m)$. However, the perceived value of the car is $\hat{V} = V(\hat{m})$, which takes a linear form

$$\hat{V} = V(\hat{m}) = v - \alpha\hat{m}, \tag{19.87}$$

where $v \geq 0$ is the component of valuation that is independent of the mileage, and $\alpha \geq 0$ is a measure of how quickly the car is perceived to depreciate. As consumers become maximally inattentive, $\theta = 1$, $V(\hat{m})$ becomes a step function.

Consider the change in perceived value for any two odometer readings $m_1$ and $m_2$ such that $m_1 - m_2 = 1$. Two cases arise.

(i) $m_1, m_2$ do not straddle two different thresholds, e.g., $m_1 = 43{,}246$ and $m_2 = 43{,}245$.
(ii) $m_1, m_2$ straddle different thresholds, e.g., $m_1 = 50{,}000$ and $m_2 = 49{,}999$.

For these two cases, (i) and (ii), a simple calculation, using (19.86), shows that

$$\hat{m}_1 - \hat{m}_2 = \begin{cases} 1-\theta & \text{Case (i)} \\ 1+\theta \times 9999 & \text{Case (ii)} \end{cases}. \tag{19.88}$$

Using (19.87) and (19.88), we get $V(\hat{m}_1) - V(\hat{m}_2) = -\alpha\left(\hat{m}_1 - \hat{m}_2\right)$, thus,

$$V(\hat{m}_1) - V(\hat{m}_2) = \begin{cases} -\alpha(1-\theta) & \text{Case (i)} \\ -\alpha(1+\theta \times 9999) & \text{Case (ii)} \end{cases}. \tag{19.89}$$

Clearly, there is a discontinuous drop in the perceived value of the car as we move from Case (i) to Case (ii). Consider Case (ii). From the second row in (19.89), we can write $V(\hat{m}_1) - V(\hat{m}_2) = -\alpha\left[(m_1 - m_2) + \theta \times (10000 - (m_1 - m_2))\right]$. Thus,

$$\lim_{m_2 \to m_1}\left[V(\hat{m}_1) - V(\hat{m}_2)\right] = -\alpha\theta \times 10000,$$

gives the discontinuous drop in car value at a threshold.

These predictions are nicely summarized in the self-explanatory Figure 19.12 for $50{,}000 \leq m \leq 100{,}000$. Since we have assumed a constant rate of depreciation, embodied in a constant value of $\alpha$, the discontinuous downward jumps at each threshold are of the same size. However, comparing cars of different makes or models that have different rates of depreciation, more rapidly depreciating cars (higher $\alpha$) will have larger discontinuous jumps at the thresholds.

The data for testing these predictions comes from wholesale markets in used cars. A simple theoretical model can be sketched that the reader might wish to formalize further. Suppose that

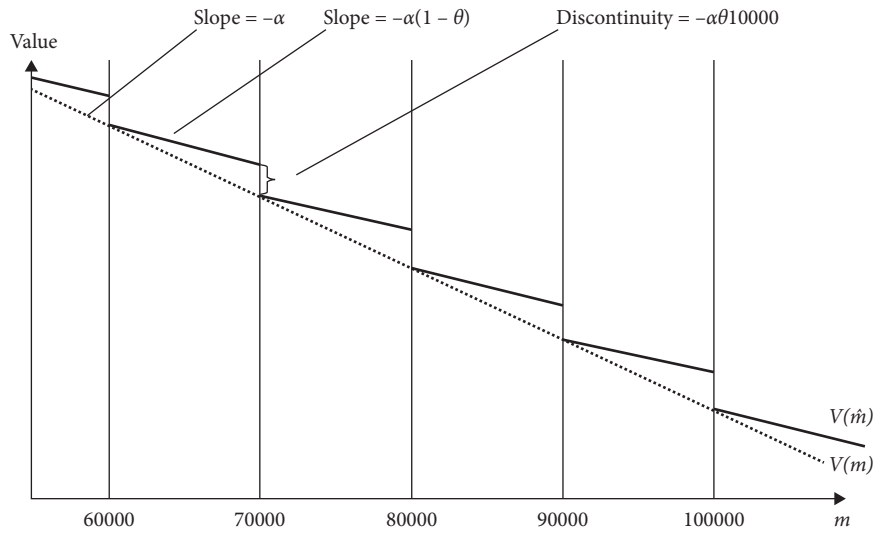


**Figure 19.12** Predictions of a model of limited attention on used car values.
Source: Lacetera et al. (2012) with permission from the American Economic Association.

there are $N$ consumers with unit demands for cars and whose perceived value for a car is given in (19.87). Cars are sold in a competitive retail market by dealers who, in turn, purchase their cars from auctions. The auctions are either second price auctions or ascending-bid English auctions. There are $M \leq N$ cars in the auctions; each car displays an odometer mileage, and its reserve price is zero. Since the market for dealers is competitive, they make no profits. Hence, in the auction market, they bid as if they had the valuations of the buyers in (19.87). Given these assumptions, the auction produces an efficient outcome, so for a car of any mileage, $m$, the auction price equals $V(\hat{m})$.

Figure 19.13 shows the raw prices of cars based on data for 22 million cars sold during the sample period, 2002–2008, at one of the largest wholesale car price auctions in the US. In the figure, the mileage is rounded to the nearest 500 (so, e.g., 42,315 is rounded to 42,500). The vertical lines in the figure represent each of the 10,000 mile thresholds from 0 to 125,000 miles. Average prices decline with mileage. Within two consecutive threshold values, the decrease in price is continuous. However, at each threshold value, there is a discontinuous downward jump in price, as predicted by models of limited attention.

Repeating the same analysis by netting out the effects of various fixed effects, such as the model, make, year, body style of cars, and plotting the price residuals gives a result very similar to Figure 19.13; the price plot is even smoother in this case but the price discontinuities survive. There are spikes in the volume of cars brought to the auctions with odometer readings just short of the 10,000 mile thresholds; this is most noticeable at thresholds starting at 60,000 miles. This suggests that car sellers are aware of the problem of limited attention on the part of car buyers. Thus, the left-digit bias is a feature of the demand side of the market rather than the supply side.

Using data from German used car markets, Englmaier et al. (2014) replicate the results above about the fall in the value of used cars when the odometer hits critical threshold values. However, they also find support for an additional and plausible factor in explaining discontinuous drops in used car values, the vintage or year of registration of the car. The left-digit bias cannot be used

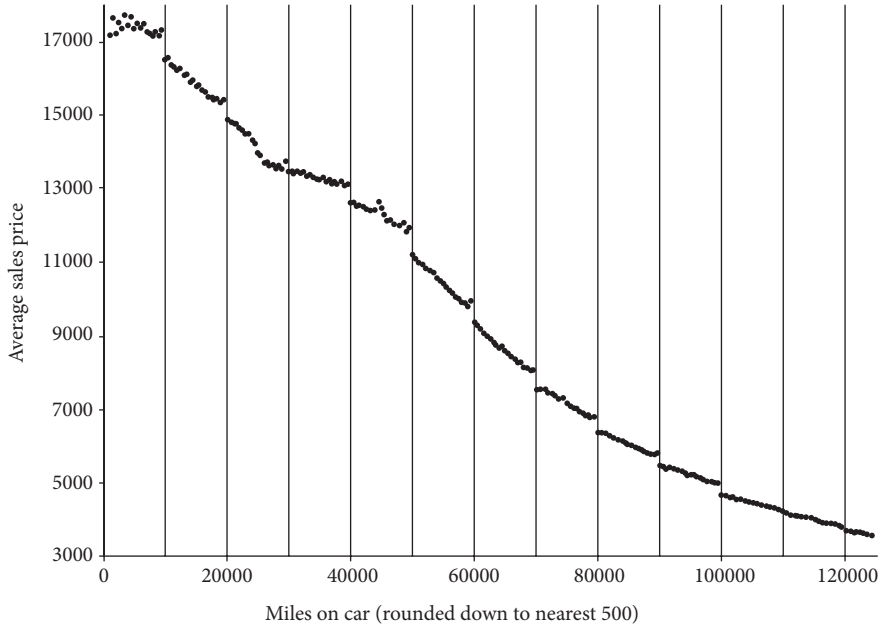


**Figure 19.13** Raw prices of cars based on data for 22 million cars sold during the sample period, 2002–2008. Source: Lacetera et al. (2012) with permission from the American Economic Association.

to account for the vintage effect, so they propose a slightly different theoretical model to account for both effects

## 19.17.4 *A theoretical framework*

In neoclassical economics, an individual may be expected to take account of a very large number of variables in solving a problem. However, cognitive limitations may allow limited attention to some factors and greater attention to more salient factors; variables given limited attention might be substituted by default values. The choice of which factors to treat as salient may be an endogenous one. Once the individual decides on the set of salient factors, he proceeds as in neoclassical models in economics. Models with these features, where the decision is made in a manner that is consistent with neoclassical models of rationality, are sometimes known as *rational inattention models*. Our main focus in this section is on the *sparse max operator* model of Gabaix (2014) that we consider to be psychologically better grounded, and more tractable. We offer some brief comments on other models at the end.

Consider the classical problem of maximizing utility $\max_{x \in X} u(x, \theta)$, where $X \subset R^k$ is the $k$ dimensional real domain of choices under consideration, and $\theta$ is a vector of parameters drawn from some set of parameters $\Theta \subset R^n$. We wish to adapt this problem to reflect limited attention by defining a *sparse max* operation,

$$s \max_{x \in X} u(x, \theta). \tag{19.90}$$

Denote $\theta = 0$ as the *default parameter* and the corresponding optimal action $x^d = \max_{x \in X} u(x, 0)$ as the *default action*. We assume that $u$ is concave in $x$ and twice continuously differentiable around $(x^d, 0)$. Suppose that the parameter $\theta$ is a linear combination of $n$ other parameters $\theta_1, \theta_2, \ldots, \theta_n$.

$$\theta = \sum\nolimits_{i=1}^{n} \mu_i \theta_i, \text{ and } \forall i, \mu_i \in R, \theta_i \in \Theta.$$

Limited attention is captured as follows. The perceived value, $\hat{\theta}_i$, of the parameter $\theta_i$ is given by

$$\hat{\theta}_i = a_i \theta_i, i = 1, \ldots, n, \tag{19.91}$$

where $a_i \in [0,1]$ is the attention given to $\theta_i$ and $a = (a_1, a_2, \ldots, a_n)^T \in [0,1]^n$ is known as the *attention vector*; superscript $T$ denotes the transpose. $a_i = 1$ corresponds to the case of full attention in neoclassical economics. At the other extreme, when $a_i = 0$, the parameter $\theta_i$ is given no attention. Once the optimal attention vector $a^* = (a_1^*, a_2^*, \ldots, a_n^*)^T$ is endogenously determined, we can calculate the perceived or sparse parameter

$$\hat{\theta}\left(a^*\right) = \sum\nolimits_{i=1}^{n} \mu_i \hat{\theta}_i = \sum\nolimits_{i=1}^{n} \mu_i a_i^* \theta_i.$$

The individual then solves the following problem

$$x^* \in \arg\max_{x \in X} u(x, \hat{\theta}(a^*)). \tag{19.92}$$

The first order condition is $u_x(x^*(\hat{\theta}\,(a^*)), \hat{\theta}\,(a^*)) = 0$, where $x^*(\hat{\theta}\,(a^*))$ is the optimal solution. Substituting the optimal solution, $x^*(\hat{\theta}\,(a^*))$, in the direct utility function, $u$, we get the indirect utility function

$$v(a) = u(x^*(\hat{\theta}(a^*)), \theta). \tag{19.93}$$

Using the implicit function theorem on the first order condition and evaluating derivatives at $(x, \theta) = (x^d, 0)$, we get

$$x_{\theta_i} \equiv \frac{\partial x}{\partial \theta_i} \mid_{(x^d,0)} = -\frac{u_{x\theta_i}}{u_{xx}} \mid_{(x^d,0)}. \tag{19.94}$$

Thus, $x_{\theta_i}$ is the change in the optimal action when $\theta_i$ increases by one unit at the default option.

Consider, henceforth, the special case of quadratic utility and $x \in \mathbb{R}$.

$$u(x, \hat{\theta}) = \frac{-1}{2}\left(x - \sum\nolimits_{i=1}^{n} \mu_i a_i^* \theta_i\right)^2. \tag{19.95}$$

In this case, the optimal solution is

$$x^* = \sum\nolimits_{i=1}^{n} \mu_i a_i^* \theta_i. \tag{19.96}$$

We now model the optimal determination of attention. Suppose that attention is cognitively costly and the cost function of attention to $\theta_i$ is given by

$$C(a_i) = ca_i^{\alpha}, c > 0, \alpha \geq 0. \tag{19.97}$$

When $\alpha = 0$, there is a fixed cost of attention equal to $c$. In classical game theory, attention is costless, so $c = 0$. Suppose that the vector $\theta$ is drawn from some distribution $g(\theta)$ such that the means of the marginal distributions are zero ($E\theta_i = 0$), the standard deviations are $\sigma_i = E\left[\theta_i^2\right]^{1/2}$, and the covariances are given by $\sigma_{ij} = E\left[\theta_i\theta_j\right]$. The total cost of attention is given by

$$C(a) = \sum\nolimits_i C(a_i).$$

The individual wishes to choose the attention vector, $a$, in order to maximize the expected value of indirect utility in (19.93), net of the cost of effort given in (19.97). Thus, the individual's problem is

$$\max_{a\in[0,1]^n} Ev(a) - C(a).$$

To simplify the problem we take a Taylor expansion of $v(a)$ around $a \equiv i = (1, 1, \ldots, 1)^T$, where $i$ is a vector of ones.

$$v(a) \approx v(i) + Dv(i)\,(a - i) + \frac{1}{2}\,(a - i)^T D^2 v(i)\,(a - i), \tag{19.98}$$

where $Dv(i)$ is the matrix of first order partial derivatives of $v$ and $D^2v(i)$ is the matrix of second order partial derivatives. $(a - i)$ is a row vector with the *jth* element $a_j - 1$. Using the definition of indirect utility in (19.93), $v(a) = u(x^*(\hat{\theta}\,(a)), \theta)$, and (19.91), differentiating once with respect to $a_i$ we get $v_{a_i} = u_x x_{\hat{\theta}_i}\theta_i$. Since $E\theta_i = 0$ we have $ED v(i)\,(a - i) = 0$ (this is the zero vector). Differentiating $v_{a_i}$ with respect to $a_j$ we get $v_{a_i a_j} = x_{\hat{\theta}_i} u_{xx} x_{\hat{\theta}_j}\theta_i\theta_j$ (recall that utility is quadratic, (19.95), and the optimal solution is (19.96), so the cross partials are zero). Taking expected values we get $Ev_{a_i a_j} = E[\theta_i\theta_j]x_{\hat{\theta}_i} u_{xx} x_{\hat{\theta}_j} = \sigma_{ij} x_{\hat{\theta}_i} u_{xx} x_{\hat{\theta}_j}$. Using these results, and taking the expected value in (19.98) we get

$$E\left[v(a) - v(i)\right] \approx \frac{-1}{2}\,(a - i)^T \Lambda\,(a - i),$$

where $\Lambda$ is the *cost of inattention matrix* such that

$$\Lambda_{ij} = -\sigma_{ij} x_{\hat{\theta}_i} u_{xx} x_{\hat{\theta}_j}. \tag{19.99}$$

Thus, an individual who uses the sparse max operator will choose the vector of attention $a$ to

$$\underset{a\in[0,1]^n}{\text{maximize}}\ \frac{-1}{2}\,(a - i)^T \Lambda\,(a - i) - C(a). \tag{19.100}$$

We summarize this in a formal definition.

**Definition 19.6** *(Unconstrained sparse max operator): The individual undertakes the following two-step optimization exercise.*

*(1) Choose optimal attention, $a^*$, to solve (19.100). This gives the sparse representation of $\theta_i$, $\hat{\theta}_i = a_i^* \theta_i$, $i = 1, \ldots, n$.*
*(2) The individual then solves the optimization problem in (19.92) giving rise to the vector of optimal choices, $x^*$.*

Consider a simple one-dimensional example, $k = 1$, $n = 1$, quadratic utility, and $c = 1$. In this case, the analog of (19.100) is

$$\max_{a \in [0,1]} \frac{-1}{2} (a-1)^2 \sigma^2 - a^\alpha. \tag{19.101}$$

The optimal values of $a$ for three different values of $\alpha = 0, 1, 2$, denoted, respectively, by $a_0^*$, $a_1^*$, and $a_2^*$, are given by

$$a_0^* = \begin{cases} 0 & \text{if} \quad \sigma^2 < 2 \\ 1 & \text{if} \quad \sigma^2 > 2 \end{cases}.$$

$$a_1^* = \begin{cases} 0 & \text{if} \quad \sigma^2 < 1 \\ 1 - \frac{1}{\sigma^2} & \text{if} \quad \sigma^2 \geq 1 \end{cases}.$$

$$a_2^* = \frac{\sigma^2}{2 + \sigma^2}.$$

A plot of the optimal values against $\sigma^2$ (in bold) is shown in Figure 19.14; $a_0^*$ (left panel), $a_1^*$ (middle panel), and $a_2^*$ (right panel). It is obvious that $a_0^*$, $a_1^*$ induce sparsity. In both cases, there is a range of values for which the optimal attention is 0; $a_1^*$ is continuous but $a_0^*$ is not continuous at $\sigma^2 = 2$. In these cases, the individual is rationally inattentive in the sense that he takes account of the costs of attention as well as the stochastic structure of the problem. However, $a_2^*$ never induces sparsity. We leave the proof of the next proposition for the reader.

**Proposition 19.7** *(Gabaix, 2014): Suppose that $\sigma_{ij} = 0$ for all $i \neq j$. Then the optimal attention, $a_i^*$, is increasing in $\sigma_i^2$, $x_{\theta_i}$, and decreasing in c.*

Optimal attention is increasing in $\sigma_i^2$ and $x_{\theta_i}$ because these factors increase the loss in utility from imperfect attention (see (19.99)). An increase in $c$ increases the cost of paying attention (see (19.100)), hence, it reduces optimal attention.

Gabaix (2014) offers plausible psychological interpretations of his model. For instance, Step 1 in Definition 19.6 may be considered to be a System 1 operation that decides how much attention to

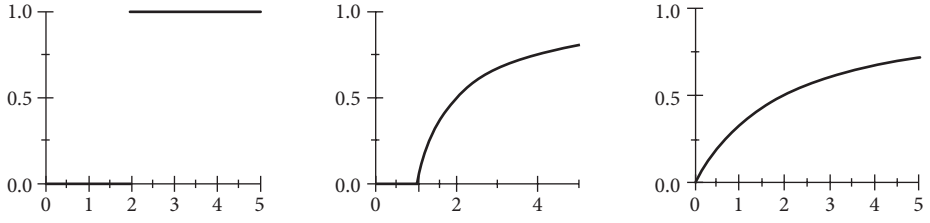


**Figure 19.14** Determination of the optimal attention parameters as a function of the variance of the underlying parameter. The horizontal axis in each of the three panels is $\sigma^2$. The vertical axes in the three panels are: $a_0^*$ (left panel), $a_1^*$ (middle panel), and $a_2^*$ (right panel).

give to a problem. Once this problem is solved, then Step 2 appears akin to a System 2 problem. The case of no attention, $a_i^* = 0$, may be considered to arise when System 1 operates under intense time pressure, and so assigns the default option $\left(x^d, 0\right)$. One may also consider this as an anchoring and adjustment model in which one begins at the default option $\left(x^d, 0\right)$ and then slowly adjusts towards the optimum $(x^*, a^*)$ although this often falls short of the full attention optimum $(x, 1)$.

Once this is accepted, the model can be extended to constrained optimization. Suppose that now the problem is to optimize, as in (19.90), subject to the constraint $f(x, \theta) \geq 0$. Define the Lagrangian

$$L = u(x, \theta) + \lambda^d f(x, \theta), \tag{19.102}$$

where $\lambda^d$ is the Lagrangian multiplier.

**Definition 19.7** *(Constrained sparse max operator; Gabaix, 2014): In the constrained problem, the sparse max operation, s* $\max_{x \in X} u(x, \theta)$*, subject to* $f(x, \theta) \geq 0$*, is defined as follows. The individual undertakes the following two-step optimization exercise.*

*(1) Choose optimal attention, a*, to solve (19.100), but instead of (19.99), using* $\Lambda_{ij} = -\sigma_{ij} x_{\hat{\theta}_i} L_{xx} x_{\hat{\theta}_j}$ *and* $x_{\hat{\theta}_j} = -\frac{L_{x\theta_i}}{L_{xx}} \mid_{(x^d, 0)}$ *where* $x^d$ *solves* $\arg\max_x L(x, 0)$*. This gives the sparse representation of* $\theta_i$*,* $\hat{\theta}_i = a_i^* \theta_i$*,* $i = 1, \ldots, n$*.*

*(2) Define the function* $x(\lambda) = \arg\max_x \left[ u(x, \hat{\theta}) + \lambda f(x, \hat{\theta}) \right]$*. Now solve the problem*

$$\lambda^* \in \arg\max_{\lambda} u(x(\lambda), \hat{\theta})$$

*subject to*

$$f(x(\lambda), \hat{\theta}) \geq 0.$$

*The solution to this optimization problem gives rise to the vector of optimal choices,* $x^*(\lambda^*)$*.*

Gabaix (2014) uses this framework to re-derive basic results in consumer theory in microeconomics and in competitive equilibrium. Some of these results include the following. Individuals optimally exhibit money illusion. This can arise because consumers are inattentive to an increase in price but attentive to an increase in their incomes. The Slutsky matrix may not be symmetric under limited attention, and a behavioral analog of the Edgeworth box is derived. The weak axiom of revealed preference may also not hold.

Overall, this is an important contribution that is better psychologically founded than many of the other theoretical contributions. But, in the spirit of the approach in the rest of this book, it must await stringent testing before it gains greater acceptance. For that it must address questions such as the following. Is imposing a stochastic structure on the unknown parameters of the utility function the most attractive approach to modeling limited attention? What are the experimental/field analogs of $\sigma_i^2$, $x_{\theta_i}$, and $c$? In several of the examples that we considered above, such as a fall in demand when sales taxes (known to people) are made salient, or changes in the pattern of Internet shopping when the less-salient delivery charges (known to people) are varied, the observed environments appear to be non-stochastic. How should we model limited attention in these cases? One may expect much progress to be made in future years on these questions.

In a popular class of models of rational inattention, information is modeled as a reduction in uncertainty. Uncertainty, in turn, is modeled as a reduction in *entropy*. Finally, an upper limit on the flow of information is imposed (Sims, 1998, 2003, 2006). Consider the simple exposition of the basics of this approach in Wiederholt (2010). Suppose that a random variable $Y$ is distributed normally with mean $\mu_Y$ and variance $\sigma_Y^2$. The *unconditional entropy* of $Y$ is given by

$$S(Y) = \frac{1}{2}\log_2(2\pi e\sigma_Y^2). \tag{19.103}$$

The *conditional entropy* of $Y$, conditional on another normally distributed variable $Z$, is given by

$$S(Y \mid Z) = \frac{1}{2}\log_2(2\pi e\sigma_{Y|Z}^2), \tag{19.104}$$

where $\sigma_{Y|Z}^2$ is the conditional variance. A measure of the information that the random variable $Z$ contains about $Y$ is given by

$$\Delta S = S(Y) - S(Y \mid Z).$$

Observing $Z$ reduces uncertainty about the variable $Y$. In this literature, limited attention is modeled as an upper bound, $b > 0$, on the information flow

$$\Delta S \leq b. \tag{19.105}$$

Using (19.103), (19.104) in (19.105) we get $\log_2\left(\frac{2\pi e\sigma_Y^2}{2\pi e\sigma_{Y|Z}^2}\right) \leq 2b$, or

$$\frac{\sigma_Y^2}{\sigma_{Y|Z}^2} \leq 2^{2b}. \tag{19.106}$$

We may think of $Z$ as a noisy signal of $Y$ so that $Z = Y + \varepsilon$ where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. In this case, (19.106) can be written as

$$\frac{\sigma_Y^2}{\sigma_\varepsilon^2} \leq 2^{2b} - 1. \tag{19.107}$$

Thus, in this literature, limited attention imposes bounds on the signal to noise ratio in a purely statistical manner. These ideas have been used successfully in physical communication channels such as characterizing the rate of information flow in modems and Internet connections. It is an open question if these statistical techniques also describe similar information flows for humans. In rational inattention models, one minimizes two kinds of costs: departures of decisions from an optimum and the cost of paying attention, subject to a constraint such as (19.107). One criticism of the earlier models in this tradition is that they cannot model source-dependent inattention but recent work has attempted to address this criticism (Maćkowiak and Wiederholt, 2009; Woodford, 2012). Gabaix (2014, p. 1696) argues that these models still lack the tractability and generality of the sparse max operator.

## 19.18 Do experts exhibit biases?

Does market experience eliminate behavioral anomalies? The evidence gives a strongly negative answer, although experience may lessen some anomalies. In this section, we look at expert-bias in the context of judgment heuristics.

Existing empirical work suggests that experts are not immune to biases. Tetlock (2002) looks at the prediction of experts of political events. His results are sufficiently striking to justify a detailed quote.

> Across all seven domains, experts were only slightly more accurate than one would expect from chance . . . . Most experts thought they knew more than they did. Across all predictions that were elicited, experts who assigned confidence estimates of 80% or higher were correct only 45% of the time . . . Expertise thus may not translate into predictive accuracy but it does translate into the ability to generate explanations for predictions that experts themselves find so compelling that the result is massive over-confidence.

Most evidence shows that experts are typically more overconfident relative to lay people who lack experience (Heath and Tversky, 1991; Frascara, 1999; Glaser et al., 2007a,b) and that overconfidence increases with experience (Kirchler and Maciejovsky, 2002). The predicted distributions of stock market returns by senior finance professionals, in a large dataset, are too narrow, and realized returns are within their 80% confidence intervals, only 36% of the time (Ben-David et al., 2013).

In this chapter, we have reported many empirical findings that demonstrate several kinds of expert-bias that we quickly summarize now. For surveys, see, for instance, Gilovich et al. (2002), Kahneman (2003), Kahneman (2011), and Tetlock (2006). Tetlock (2010) is a useful response to alternative critical views. Mathematical psychologists exhibit the law of small numbers (Tversky and Kahneman, 1973). The affect heuristic determines experts' perceived riskiness of various hazardous substances (Slovic et al., 1999) and the expert-assigned likelihood of offending of recently discharged mental patients (Slovic et al., 2000). Clinical psychologists underweight the base rate relative to a Bayesian decision maker (Meehl and Rosen, 1955). Expertise does not shield decision makers from the hindsight bias, as shown in a meta-study (Guilbault et al., 2004). Traders in a large investment bank, in both its international branches, were found to be hindsight-biased (Biais and Weber, 2009).

People are often willing to pay a premium for expert services that are mistakenly perceived to be competent (Powdthavee and Riyanto, 2015). Evidence supports the important role of anchoring on a list price by estate agents (Northcraft and Neale, 1987) and in legal judgment (Chapman and Bornstein, 1996; Englich and Mussweiler, 2001; Englich et al., 2006). The false consensus effect is exhibited by judges (Solan et al., 2008). Finance professionals also exhibit a false consensus effect by imputing on others, their own risk preferences (Roth and Voskort, 2014).

Experts, such as physicians and World Bank staff, exhibit framing effects, often of a similar magnitude to those observed with student populations. When presented with identical options, expert responses can flip depending on whether the outcomes are framed as gains or losses relative to a reference point (McNeil et al., 1982; Kahneman and Tversky, 1984; WDR, 2015). There is evidence of limited attention in wholesale car markets by wholesale car dealers (Lacetera et al., 2012).

In the experiments of Redelmeier and Shafir (1995), physicians were given a case study in which a patient suffers chronic hip pain. The physician, in the case study, is on the verge of referring the

case to an orthopedic consultant. At this point, the physician in the case study learns new information about drugs that the patient has not tried. In one treatment, physician-subjects learn that the patient has not yet tried Ibuprofen, and in the second treatment they learn that the patient had not yet tried Ibuprofen and Piroxicam. In the first treatment there were two choices: Ibuprofen+referral and plain referral. In the second treatment, in addition to these two choices, an extra choice was added: Piroxican+referral. In the second treatment, despite more choices (three instead of two), 72% of the physician-subjects chose plain referral as compared to 53% in the first treatment.

Being subject to judgment heuristics and framing effects, experts can make systematic errors just like anyone else. Thus, market experience does not eliminate non-standard behaviors. But these are not the only reasons that experts exhibit potential biases. Moskowitz and Wertheim (2011) argue that the home ground advantage in sports arises not just from insider-knowledge of home players but also from expert-biases. The experts, in this case, are the match referees. Even if the referees are completely impartial and honest, they are subject to a natural human tendency to avoid the displeasure of those in the proximity. Thus, their discretionary authority typically tends to be exercised in favor of the home team, as for instance, in adding discretionary injury time minutes at the end of a football game. The idea of "home ground advantage" in sports is also used to explain the behavior of experts in the 2008 financial crises by Barth et al. (2013). Many financial experts held senior private sector positions in the financial sector before holding similar senior positions in the US Fed. The home team, in this case, becomes the private financial sector and the away team, the dispersed individual investors in the financial sector, whose interests were insufficiently protected.

In an admirable piece of work, the WDR (2015) documents several kinds of biases among its professional staff. These include confirmation bias, susceptibility to sunk costs, and the influence of framing. The WDR (2015, p. 18) is candid in its assessment of expert-bias: "This finding suggests that development professionals may assume that poor individuals may be less autonomous, less responsible, less hopeful, and less knowledgeable than they in fact are." The WDR also suggests potential solutions to the problem of expert-bias. These include *dogfooding* (experts signing up and playing their own programs for real) and *red teaming* (having an adversarial outside team test that tests the proposals).