

Centre de Referència en Economia Analítica

Barcelona Economics Working Paper Series

Working Paper n° 39

**Is confidence in Decisions Related to Feedback?
Evidence –and lack of evidence- from Random
Samples of Real-World Behavior**

Robin Hogarth

November 2003

**Is confidence in decisions related to feedback? Evidence – and lack of evidence –
from random samples of real-world behavior ***

Robin M. Hogarth*

ICREA & Universitat Pompeu Fabra, Barcelona

Barcelona Economics WP no. 39

To appear in Klaus Fiedler and Peter Juslin (Eds), *In the beginning there is a sample: Information sampling as a key to understand adaptive cognition*. Cambridge, UK: Cambridge University Press.

November 2003

Abstract

Confidence in decision making is an important dimension of managerial behavior. However, what is the relation between confidence, on the one hand, and the fact of receiving or expecting to receive feedback on decisions taken, on the other hand? To explore this and related issues in the context of everyday decision making, use was made of the ESM (Experience Sampling Method) to sample decisions taken by undergraduates and business executives. For several days, participants received 4 or 5 SMS messages daily (on their mobile telephones) at random moments at which point they completed brief questionnaires about their current decision making activities. Issues considered here include differences between the types of decisions faced by the two groups, their structure, feedback (received and expected), and confidence in decisions taken as well as in the validity of feedback. No relation was found between confidence in decisions and whether participants received or expected to receive feedback on those decisions. In addition, although participants are clearly aware that feedback can provide both “confirming” and “disconfirming” evidence, their ability to specify appropriate feedback is imperfect. Finally, difficulties experienced in using the ESM are discussed as are possibilities for further research using this methodology.

Key words: Decision making, learning, confidence

JEL Codes: M10

A central theme of this book is how people learn from samples. In this chapter, we look at this issue from two viewpoints. First, we question the manner in which we – as scientists – sample the environments of the experimental participants who engage in the judgment and decision making tasks that we study. To what extent are these samples representative of the natural decision making ecology that our participants face in their everyday lives? Second, by actually sampling people’s activities in their natural ecologies, we seek to characterize how they experience these environments. In particular, we investigate how one feature of environments (the presence or absence of feedback) affects inferences (the confidence people express in their decisions).

The first question is, of course, not new. For many years, psychologists have been concerned about how to generalize behavior from experimental evidence (see, e.g., Brunswik, 1956; Chapanis, 1961; 1967; Cronbach, 1975; Ebbesen & Konečni, 1980; Hammond, 1978; Hogarth, 1986; Lipshitz, Klein, Orasanu, & Salas, 2001). I do not propose to add to this debate. Instead, the contribution is to demonstrate how the use of readily available technology can greatly facilitate random sampling of decision behavior outside the psychological laboratory. Indeed, as I shall argue below, obtaining appropriate samples of human decision behavior is not as difficult as might be imagined by researchers trained within experimental paradigms.

The intellectual stimulus for this work was a study reported by Egon Brunswik in 1944 and the development of the *experience sampling method* or ESM by Csikszentmihalyi and others (see Brunswik, 1944; Csikszentmihalyi & Larson, 1987; Hurlburt, 1997). Briefly stated, this chapter reports a study of the decision behavior of two groups of people, business executives and students. For several days, participants received 4 or 5 SMS messages (on their mobile telephones) at random

moments at which point they completed brief questionnaires about their current decision making activities. With this methodology, therefore, it was possible to achieve random samples of each participant's decision behavior although, of course, no claim can be made that the participants themselves were anything other than "convenience" samples of their respective populations, that is, business executives and students.

The samples of decisions obtained by the ESM can be used to infer the characteristics of people's decision environments. This, in turn, can be used to address our second issue that deals with how such characteristics affect the inferences that people make. To do this, we focus on the issue of confidence and ask how one aspect of environments – namely the presence or absence of feedback – affects the confidence that people express in their decisions.

The issue of whether people exhibit appropriate confidence in their judgments has attracted much attention in the decision making literature. Here, the key concept has been calibration, that is, do people's assessments of uncertainty match empirical relative frequencies? For example, do events that are assessed subjectively as occurring, say, 65% of times actually occur 65% of times, and so on? Early findings suggested that people are overconfident, that is, their assessments of probabilities of target events are systematically higher than empirical relative frequencies (see, e.g., Lichtenstein, Fischhoff, & Phillips, 1982). However, these experimental findings have been challenged as researchers have shown, inter alia, differential effects of some realistic task environments (Murphy & Winkler, 1984), framing questions in terms of frequencies as opposed to probabilities (Gigerenzer, Hoffrage, & Kleinbölting, 1991), "hard" and "easy" questions (Harvey, 1997; Juslin, 1993), and

whether confidence is expressed in the form of confidence intervals or beliefs about binary choices (Klayman, Soll, González-Vallejo, & Barlas, 1999).

More importantly, Gigerenzer et al. (1991) adopted a Brunswikian perspective and demonstrated how the representative sampling of questions from people's natural environments led to judgments of confidence that were well calibrated. In addition, Juslin and his colleagues have made a careful examination of how many other experimental tasks have been sampled and have demonstrated that non-representative sampling of items could account for much overconfidence (Juslin, 1994; Juslin, Winman, & Olsson, 2000). In other words, if research on this topic had sought to sample tasks in the ecologically valid manner advocated by Brunswik (1956), "overconfidence" might not be considered such an important issue in behavioral decision making as it is today (see, e.g., Russo & Schoemaker, 2002).

The calibration paradigm tests people's ability to express their confidence in probabilistic form for particular events or classes of events. However, it does not address the more general issue of whether people feel confident in their everyday decision making activities and, if so, whether these feelings are justified. Nonetheless, in terms of behavior, such feelings are important. They affect how individuals feel and act; and they can also influence others. In the workplace, for example, the confidence expressed by one party (e.g., a boss or a salesman) may determine whether a decision is implemented (e.g., to undertake an action or buy a product).

Some 25 years ago, Einhorn and I (Einhorn & Hogarth, 1978) addressed the issue of confidence in judgment from a theoretical perspective. We noted that the structure of many decision making tasks in the real world is such that people either do not receive feedback on their decisions or, if feedback is received, it can be

misleading. For example, consider hiring decisions made by managers. By observing how well new employees perform in their jobs, managers clearly receive feedback on their hiring decisions. However, this feedback is incomplete because, by hiring some applicants but not others, there is no feedback concerning the performance of rejected candidates. In addition, the very act of hiring may engender self-fulfilling prophecies with respect to “successful” candidates. The theoretical implication is that people’s general confidence in their decision making abilities can be continually reinforced by the fact that many of their actions involve positively biased or even no feedback (with the latter being interpreted as “no news is good news”). Subsequent experiments by Camerer (1981) and Schwartz (1982) supported this analysis. (For related ideas and evidence, see Fiedler, 2000).

Einhorn and I subsequently refined these ideas by characterizing environments in which erroneous beliefs might be created and maintained by distorted or even missing feedback. Einhorn (1980) referred to outcome irrelevant learning structures or OILS. More recently (Hogarth, 2001), in reviewing how people develop “good” and “bad” intuitions, I introduced the notion that the environments in which people acquire such knowledge can be classified as “kind” or “wicked” according to whether the feedback they receive is or is not veridical (i.e., accurate).¹

This chapter addresses the relation between decision making and confidence in people’s natural decision making environments. Specifically, is there a relation between the confidence that people express in their decisions and the feedback that they either receive or expect to receive? The data reported are generated by a novel methodology for research on decision making inspired by Brunswik’s (1944) study of “size constancy.”² (See also Hoffrage & Hertwig, this volume.) In this study, Brunswik arranged for a person (a student) to be followed by an associate for

extensive periods of time in her natural environment (the University of California at Berkeley) over a period of four weeks. The participant was instructed to behave in her normal fashion but was asked – at irregular moments – by the associate to estimate the sizes of objects in her visual field when interrupted as well as the distance from the objects. The associate then measured the objects and distances. Note, that in this study the “experimenter” did not manipulate or choose the objects about which the participant made judgments, that is, the “experimental tasks.” Instead, these were selected in a manner such that the objects of the participant’s judgments constituted a random sample of tasks in her natural environment. Thus, although this study involved but a single participant, valid inferences could be made about the “size constancy” that she exhibited.

Briefly stated, Brunswik’s point was that the environments of psychological experiments distort reality because participants are asked to act or make judgments in situations that have been artificially created by manipulating variables in some orthogonal, factorial manner. Instead, Brunswik argued, many variables in the real world are correlated to some extent and people learn to deal with a variety of situations as opposed to isolated incidents that have been created by an experimenter. Thus, if you want to know how people “do something” in the real world, you need to sample situations from the environments (or real worlds) in which they actually live.

This chapter exploits the ESM to explore issues of confidence in decisions as well as the kind of feedback people receive and expect on the decisions they take. Specifically, studies were conducted with two distinct populations, business executives and students. However, to gain more insight into estimates of behavior generated by the ESM, a retrospective questionnaire was also administered in which another group of student respondents was asked to summarize aspects of their

decision making activities over the two preceding weeks.³ This requested aggregate level estimates of characteristics of decision making activities that corresponded to the questions that participants in the ESM studies answered for all the specific decisions on which they reported.

The retrospective questionnaire was motivated by methodological and substantive concerns. First, there is evidence that retrospective reports of events differ from data concerning the same events collected at the time the events actually occurred (see, e.g., Ericsson & Simon, 1984; McFarlane, Martin, & Williams, 1988). To what extent, therefore, would estimates of behavior obtained by ESM differ from retrospective reports? Moreover, in what way would they differ? Second, within the calibration paradigm (see above), people have been shown to exhibit less confidence in judgments when these are viewed within the framework of a series of judgments as opposed to considering each judgment separately (Gigerenzer, Hoffrage, & Kleinbölting, 1991). Does this finding generalize to feelings that people have about confidence in the decisions they take (i.e., retrospective reports vs. confidence expressed when deciding)?

The chapter is organized as follows. I first describe the participants in the study and the methodology. Specific questions centered on the kinds of decisions taken on an ongoing basis; the extent to which these decisions are repeated frequently or are unique; confidence expressed that decisions are “right;” whether people receive or expect to receive feedback on the appropriateness of their decisions; the kind of feedback and its timing relative to the moment of decision; and confidence as to whether the feedback is or will be appropriate.

Next, I present results of the study. To summarize, people – both students and executives – express considerable confidence in the decisions they take on a daily

basis even though they do not receive or expect to receive feedback for some 40% of these decisions. They are also quite confident in the quality of feedback (when received) and are well aware that this can lead to disconfirming the validity of their actions. As could be expected, executives and students differ in the kinds of decisions they face, both in terms of content and structure. For student populations, the data collected retrospectively differs significantly from that collected by the ESM. In particular, reported confidence in decisions as well as relative presence of feedback are larger when estimated in retrospect than from data gathered concurrently.

Finally, I discuss the results of the study in terms of both methodology and substance. As to the former, the ESM is clearly a useful adjunct to a decision researcher's toolbox and will become even more powerful when used in conjunction with more traditional methodologies. In terms of substance, the study emphasizes the importance of characterizing habitual behavior or routines that affect the cumulative consequences of the many decisions we take each day.

Participants

Three groups of participants participated in the ESM studies. Two groups were composed of business executives. The third group consisted of students at Universitat Pompeu Fabra in Barcelona, Spain. A further group of students at the same university completed the retrospective questionnaire concerning decisions made in the two preceding weeks.

The executives. The first group of executives was recruited by email notices sent by the president of the local alumni club of an executive MBA program in Barcelona that requested volunteers to participate in a study organized by the author (who actually knew many of the potential participants). However, not many details of the study were announced except that participants would be required to respond to

SMS messages on their mobile telephones during a period of two weeks. In return for participation, potential respondents were promised feedback on the study. Twenty managers volunteered to participate. However, data were only received from 13. Two executives explicitly replied that they had been unable to participate (one because of technical problems with his telephone); no data – or even any responses – were received from five other executives and it can only be presumed that they experienced difficulty in complying with the study's requirements and so gave up on the task. Of the 13 executives who responded, ten were male and three were female. Their ages varied between 30 and 45. Most of the executives worked in the Barcelona area except for two who were in Madrid, one in Mallorca, and one in another European country.

Following a brief presentation by the author, the second group of executives was recruited at an executive education program at CEDEP in Fontainebleau, France. These executives were similar in age and other demographic characteristics to the first group except that they worked in France, Belgium, the Netherlands, and the United Kingdom. Responses were subsequently received from 11 of the 16 executives who had expressed an interest in the study. A difference between the first and second groups of executives was that, in the second group, the executives were only asked to participate in the study for one week, i.e., five working days. The decision to reduce the length of participation reflected the difficulties experienced by the first group of executives. As with the first group, participants were promised feedback on the study.⁴

The students. Eleven students (ten undergraduates and one graduate student) responded to advertisements placed on notice boards on campus. These stated that the task would take approximately 30 minutes each day during two weeks and that

participants would be paid 40 euros for participating. They were required to own or have access to mobile telephones capable of receiving SMS messages. On completion of the experiment, the participants were paid 40 euros and were debriefed as to the purposes of the study. The median age of the undergraduates was 20 (range of 19 to 21); the graduate student was 28; there were two males and nine females. The responses of one participant (a 19-year old male) were excluded from analysis because it was deemed that they had not been completed with sufficient care.

Respondents for the retrospective questionnaire were recruited after an undergraduate class at Universitat Pompeu Fabra. The 25 volunteer respondents had a median age of 21. There were 11 females and 14 males. They were paid 5 euros each for completing the questionnaire, a task that took between 20 and 30 minutes. Since the responses of males and females did not differ significantly, they were pooled for the purpose of analysis.

Method

The student participants were asked to keep their mobile telephones “on” from 9 am to 9 pm each day, Monday through Saturday, for a period of two weeks (12 days). They were told that, each day, they would receive approximately five SMS text messages at random moments between 9 am and 9 pm. Each time they received a message, they were asked to complete a short questionnaire in respect of the most recent “decision action” they had taken. (See below.) Messages followed a standard format, e.g., “Message #2 – Robin” would signal that this was the second message being sent on that day. The procedure for the executive participants was the same except that messages were only sent on Mondays through Fridays and between 9 am and 7 pm.

To send the messages each day, time (9 am to 9 pm or 9 am to 7 pm) was divided into 10-minute segments and random numbers were assigned to the segments. The segments corresponding to the five largest numbers were the time intervals during which the messages were sent. In a few cases, I failed to send messages and thus participants did not receive exactly five messages each day. The messages were sent over the internet using commercially available messaging services provided by mobile telephone operators. Forty-eight messages were sent to the first group of executives; and between 20 and 24 messages were sent to the second group. Sixty or 61 messages were sent to the students (not all started the study on the same day).

The first group of executives received all instructions and questionnaires in written form through email and returned either hard or electronic versions of the questionnaires. A package of all materials was given to the second group of executives including an addressed envelope to facilitate returns. English was the sole language used with the executives. (At least seven respondents were English-native speakers and all had studied or were studying management courses in English.) The student participants were given verbal instructions (in Spanish) and written instructions (in English) and provided with questionnaire forms to complete and return to the author. The questionnaires were in English but almost all student participants responded in Spanish.

Instructions asked participants to focus on what were called “decision actions” (or DAs).⁵ These were defined as “any *decision* or *judgment* that you make.” Moreover, it was emphasized that these could vary from important to trivial and that, from the viewpoint of the study, it did not matter whether the DAs involved “large” or “small” consequences. Participants were further instructed to focus on the DA that was “closest in time to the moment that you receive the SMS message.” Since it was

known beforehand that executives would not keep their mobile telephones “on” all day, instructions to this group specified “If on reconnecting your mobile telephone you find a message, please act as though the message had just been sent, i.e., assume that the message was sent just before you reconnected your telephone.” At one level, this procedure could have produced a small bias in random sampling of decisions; on the other hand, it was designed to reduce distortions due to faulty memory.

The object of the questionnaire administered to the second group of students was to obtain retrospective estimates of the aggregate features of the data that had been collected from the first group of students on a concurrent basis. Thus, the questionnaire explicitly asked respondents to limit their estimates to the two preceding weeks, specifically, Mondays through Saturdays, from 9 am to 9 pm. The questionnaire also emphasized that all decisions, including the trivial, should be considered.

To illustrate the manner in which the questions were asked, consider the question in the ESM studies about confidence in decisions. Here, participants were asked to state for each decision taken “How confident are you that you took the ‘right’ decision?” Responses were made by checking one of five levels: “Very confident,” “Confident,” “Somewhat confident,” “Not confident,” and “Not at all confident.” In the retrospective questionnaire, respondents were asked “Of all the decisions that you took, what percentages of times could you express the following levels of confidence that you had taken the ‘right’ decision?” Responses were made by assigning percentage estimates to each of the five levels listed above (using exactly the same words). Below we refer to the responses of the questionnaire study as the “student controls” to indicate that these responses represent a baseline against which the ESM student data can be assessed.⁶

Results

Responses. As noted above, valid responses for the ESM studies were received from 24 executives and 10 students. The executives provided analyzable responses for 613 of 876 messages sent, i.e., 70%. However, there was considerable variation in individual response rates from 15% to 100%. The median was 85%. One could exclude the responses of those executives who provided few responses; however, there is no reason to believe that their responses were any more or less valid than those who gave many responses. Executives who submitted few responses simply stated that they had difficulty completing all requests for data.

The response rate from the students was much higher. In total, the students gave responses to 585 of the 605 messages sent to them, i.e., 97%. The median response rate per student was 98%. Given that the methodology interrupted their daily life at unpredictable moments, this is a highly satisfactory response rate.

Across both populations, there were approximately 1,200 decisions to analyze. In the questionnaire study, all 25 participants provided usable data.

Checks on data. Participants were asked to record both the time at which they received the SMS messages and the time at which the decisions they reported took place. Deviations between the former and the time at which the messages were sent provide a check on whether messages were being received at the appropriate moments. Checks between the time messages were received and the time reported decisions took place are important because the smaller the gap between these two times, the less likely it is that the decisions reported suffer from selection biases or

distortions in memory. (Recall participants were asked to report on decisions that immediately preceded the receipt of messages.)

For the 24 executives, the median times between sending messages and the recording of these varied between 1 and 46 minutes with an overall median (by participant) of 5 minutes. Thus, the executives did not always report receiving the messages when they were sent. The main reason was that several executives were not able to maintain their mobile telephones in a ready state but only consulted them from time to time. (This was a result, for example, of air travel or attending important meetings.) For all but one of the student participants, the analogous median times were only 1 or 2 minutes. (The outlier was 10 minutes.) The students clearly received the messages when they were sent.

Given the goals of the study, the more important deviation is between the moments that messages were received and when reported decisions were taken. For the executives, individual median times varied between 0 and 35 minutes with an overall median of 14 minutes. For the individual student participants, the median deviation varied from 4 to 35 minutes with an overall median (across participants) of 9 minutes.

ESM questionnaire results. Each ESM questionnaire posed 11 questions. Two questions simply asked for the date and time. Six required quantitative responses (e.g., estimates of time or checking one of several possible answers). And for three questions, participants had to write descriptions. These involved what the participants were doing when they received the messages, a description of the most recent “decision action” that they had taken, and an explanation of the feedback (if any) that they would receive on their decisions (see also below).

The strategy for dealing with the three latter questions involved, first, accepting what participants had written at face value (i.e., if a participant wrote about a decision in a particular way, it was assumed that was how she or he had actually perceived the situation); and second, developing a coding scheme to classify responses in a consistent manner. Coding involved three tasks. One was to classify what respondents were currently doing and the domain of activity to which specific decision actions applied. Categories for classification were created after reading all questionnaire responses and examining categories used in previous ESM studies (in particular, Sjöberg & Magneberg, 1990). The second task concerned the structure of each decision. This was analyzed in terms of the number of alternatives stated or implied and whether the decision was positively or negatively framed. The third task, the explanation of feedback, was classified as to whether it could be considered confirming, disconfirming, or possibly both. The schema used for analysis is discussed in greater detail below (see results concerning feedback).

All data were coded independently by two research assistants who were ignorant of the author's expectations for the study. For each qualitative question, the assistants recorded their level of agreement and then discussed all disagreements until they reached consensus. (Where appropriate, I report the initial level of agreement between the two coders.) In what follows, the data are primarily presented in aggregate form contrasting the total responses of the executives, on the one hand, and the students, on the other. Although there are variations in responses by individuals, the mean aggregate responses (of both executives and students) are almost identical to the mean responses of individuals in each group.⁷

Finally, in the tables presenting the data, I have included responses of the "student controls." These are the responses of the 25 participants in the retrospective

questionnaire study who estimated aggregate characteristics of their decision making activities over the two preceding weeks (9 am to 9 pm, Mondays through Saturdays).

Current activities and domains of decisions. When they received messages, participants were asked to record what they were doing (their “current activity”) as well as to describe their most recent decision or “DA.” Table 22.1 presents classifications of both current activities and the domains of most recent decisions⁸ – for executives and students, separately, with the data of the “student controls” appearing on the right hand side of the student data.

Insert Table 22.1 about here

The data show, first, that the domains in which decisions are taken are closely related to current activities (as one would suspect). Second, there are both similarities and differences in what the executives and students are doing and the domains of their decisions. For both, about one-third of their activities involve their lives as business people or students (“basic occupation”) with the executives making somewhat more decisions than students in this category (36% vs. 30%). On the other hand, whereas 19% of executives’ decisions involve “professional communication” (the second most important category), the analogous figure for students is a mere 2%. Indeed, if the categories of basic occupation and professional communication are summed, it is clear that the executives are involved in many more work-related activities than the students (i.e., 55% vs. 32%).

As to further differences, “sleep, rest, and recreation” are quite high on the students’ list but low on that of the executives. In addition, if we add to this category the activities and decisions devoted to “eating and drinking,” it is clear that these

occupy more space in the lives of the students than in those of the executives (27% of decisions as opposed to 13%). Interestingly, between 10% and 15% of the decisions of both groups lie in the domain “housework, personal time and funds management.” (For other studies that describe people’s activities across time using the ESM, see Sjöberg & Magneberg, 1990.)

Pre-testing of the retrospective questionnaires revealed that respondents would probably not be able to distinguish differences in estimates of the time they had spent in different activities and the domains in which they took decisions. Thus, they were only asked to estimate percentages of time spent on different activities. The distribution across activities shown in Table 22.1 under “Activities of controls” (lower right hand column) differs from the distribution of current activities reported by the ESM students in that the former is less skewed than the latter.⁹

Action types. Decisions can be described in many different ways. Above, I reported the content or domain of decisions. Here, I consider their structure. Specifically, each decision was defined as belonging to one of three different “action types.” The first is a straightforward situation where someone decides “to do” as opposed “not to do” something. What is not done is unspecified and the description of the decision focuses only on what is to be done. As an example, consider a decision described as “to have a cup of coffee.” Note that the formulation of this decision has a *positive* focus. In the second action type, we consider the decision “not” to do something, such as “not to have a cup of coffee.” This is a *negative* focus. The third action type reflects more complex decisions with multiple alternatives. An example is “to have a cup of coffee or to have a cup of tea.” More complicated examples could include “deciding on a schedule of activities” or “making a list of priorities.” These

all involve several different possible actions where the alternative is not just the negation of a single action.

Table 22.2 reports some characteristics of the data including “action type” distinguishing between responses of students and executives. Actions with a positive focus dominate those with a negative focus for both groups of participants. Indeed, the relative lack of decisions with a negative focus is striking. As to the third action type, executives clearly see more specific alternatives than students, 36% vs. 16%. However, of particular interest here is the individual variability amongst executives. Whereas for 13 (out of 24) respondents, this figure is less than 15%, for 6 it is greater than 50% (including 3 more than 95%).

.....
Insert Table 22.2 about here
.....

Orientation. Respondents were asked whether their decisions were professional or private. As might be expected, the percentage of private responses was higher for the students than the executives. Indeed, it is surprising that the students should have such a low proportion of professional responses (24%) when so many of their activities were centered on their studies (see Table 22.1). Similarly, although the executives were mainly at work, they perceived 36% of their decisions as being private in nature.

Frequency. To what extent does decision making involve frequently occurring or new activities? Here students and executives had somewhat different responses with executives reporting somewhat less frequently occurring events. (Transforming the questionnaire responses to a 1-5 scale and testing the difference in means leads to a t-statistic of 2.04, $p < .05$). Students classified 23% of decisions as being taken

“frequently and without really thinking” and for executives this percentage was 16%. There must, of course, be some doubt as to whether these latter figures represent an accurate assessment of decisions that are taken “automatically” since, almost by definition, people may not be aware of taking such decisions (see, e.g., Hogarth, 2001). On the other hand, roughly one fifth to one quarter of all decisions fell into the categories of being done “rarely” or for “the first time.”

Confidence in the “right” decision. Overall, confidence that the “right” decision was taken was high. For both students and executives, the categories of “very confident” and “confident” were checked in almost 70% of cases and relatively few responses indicated lack of confidence. Interestingly, one respondent alone accounted for almost 50% of the students’ responses to the last two categories (“not confident” and “not at all confident”) and when this participant’s responses are omitted, the data for the students and executives are more similar.¹⁰ Below, I analyze the correlates of confidence and comment further on these findings.

Finally, the retrospective data of the student controls indicate more confidence that the “right” decisions had been taken than the responses of the students in the ESM study (76% vs. 66% when summing the “very confident” and “confident” categories). In other words, the data suggest that overall assessments, based on memory, involve a greater sense of confidence than the aggregation of estimates expressed at the time events occurred. This finding runs contrary to what would be predicted from evidence in the calibration literature where overall confidence expressed for a series of events has been found to be less than that obtained by aggregating the confidence levels expressed for each of the events (Gigerenzer, Hoffrage, & Kleinbölting, 1991).

Feedback. Table 22.3 reports data concerning feedback. In light of a pilot study, the ESM questionnaire did not explicitly use the word “feedback.” Instead, respondents were asked “Will you ever know whether you took the ‘right’ decision or do you already know?” with the response categories being “Yes” or “No.” Following “Yes” responses, participants were requested to answer three questions: (1) “When will you know (please express as the length of time between taking the DA and the moment of knowing)?” (2) “Please explain how you know or will know.” (3) “How confident can you be in this explanation?” (This latter question had five possible responses – see Table 22.3.)

.....
Insert Table 22.3 about here
.....

Overall, students and executives report that some 52% to 65% of decisions are accompanied by feedback or expected feedback. (There was considerable variation in student responses but differences are not statistically significant. In particular, two students reported lack of feedback in 85% and more of cases; the largest comparable figure for any executive was 71%.) In addition, both executives and students express high levels of confidence in the accuracy of the feedback they receive or expect to receive (see foot of Table 22.3).

Where the executives and students differ is in the timing of feedback following action. Both groups do receive some feedback immediately after taking actions (32% and 20% in less than 2 minutes). Here, however, the similarity ends. Whereas, within 45 minutes students receive feedback on 55% of occasions, the analogous figure for executives is 29%. As to feedback received after a week, the figures are 20% and 35% for students and executives, respectively. In addition, in 10% of the cases the timing of feedback for executives is indeterminate, that is, it is unclear from their

descriptions when this feedback will be received, if at all. At the individual level, 14 of the 24 executives reported that at least 40% of their feedback would only be received after one week or was indeterminate, whereas this was true of only one of the 10 students.

The student controls differ from their ESM counterparts in that they claimed that they received much more feedback (70% vs. 52%) and the distribution of this feedback was more spread out in time (e.g., the median time for the student controls to receive feedback after taking decisions was between “90 minutes to 3 hours.” For the ESM student participants, this median was between “15 and 45 minutes”). As to confidence in the feedback itself, the ESM participants express greater confidence than the student controls (74% vs. 67%).

What type of feedback do people receive and expect? Participants were asked to “explain how you know or will know” that the “right” decision had been taken. For both cases, participants’ responses were classified according to a 2 x 2 table that has often been used to describe the relations between actions or beliefs and types of feedback. This is reproduced here as Figure 22.1.

The rows of Figure 22.1 represent actions taken and their alternatives; the columns distinguish between feedback that can be thought of as, alternatively, confirming and disconfirming the “correctness” or “appropriateness” of actions taken. Cell *a* represents the conjunction of taking the action and observing/expecting confirming evidence (that the “right” decision was taken); cell *b* the conjunction of taking the action and observing evidence that it was not the right decision (i.e., disconfirming evidence); cell *c* is the conjunction of not taking the action but seeing evidence that would have confirmed that the action should have been taken; and cell *d* is the conjunction of not taking the action and evidence that would have shown it to

be inappropriate. Of course, “not taking the action” also includes taking another explicit action when alternatives have been made explicit. As seen previously in Table 22.2 above, most decisions had a simple, positive focus even though there were differences between executives and students as to the extent to which alternative actions were specified (i.e., simple negations vs. multiple alternatives).

.....
Insert Figure 22.1 and Table 22.4 about here
.....

Table 22.4 provides a classification of the types of feedback received or expected. The analysis distinguishes between feedback that has been or is being received (labeled “current”) and feedback that had not yet been received (labeled “expected”). Consistent with the fact that the executives received or expected to receive feedback later than the students (relative to when decisions were taken), the students report more current feedback than the executives (32% vs. 15%) and correspondingly, less expected feedback.

The predominant type of feedback for both students and executives – and for both time periods – involves cells *a* and *b*, either separately or as a conjunction. Indeed, there are relatively few instances that involve other cells. The classification “unobservable” covers cases where participants stated that they would receive or had received feedback but, as far as could be determined from what they had written, these simply represented beliefs for which no evidence was forthcoming. Of particular interest for expected feedback is that participants were often acutely aware that their decisions could be wrong and expressed this by describing feedback in terms of both cells *a* and *b* or cells *c* and *d*. Indeed, 56% of expected feedback of the executives refers to both cells *a* and *b* (for students this figure is 36%).

To seek insight into when participants are more likely to expect “*a* or *b* cell feedback,” I regressed the corresponding percentage responses for all ESM participants on different combinations of percentage responses to other questions as well as a dummy variable to distinguish executives and students. This analysis revealed positive significant effects for executives vs. students ($t = 2.41$), decisions taken in the domain of “basic occupation” ($t = 2.69$), multiple alternative action types ($t = 1.86$), and a negative effect for feedback received within 2 minutes of making decisions ($t = - 2.15$). The R^2 from this regression equation was 0.54, $F = 8.35$, $p < .001$. These results can be interpreted by stating that respondents were more open to the possibility that feedback could be either favorable (cell *a*) or unfavorable (cell *b*) in cases involving more complex, work-related decisions for which feedback was not expected within a short time. Moreover, this was truer for executives than students.

Finally, an attempt was made in the retrospective questionnaire to elicit estimates of the kind of feedback that respondents thought they had received or would receive. These student control data are similar to the ESM results concerning current feedback but quite different for expected feedback (see right hand side of Table 22.4).

Confidence, feedback, and time. What are the correlates of feedback? First, recall from above that whereas participants received or expected to receive feedback for some 60% of their decisions, they were either “confident” or “very confident” for almost 70% of their decisions. However, there was no relation between receiving or not receiving feedback and confidence in decisions. At the individual level, the relation between feedback and confidence in decisions was statistically significant ($\chi^2 < .05$) for only 3 of the 10 students, and for one of these respondents, the relation was negative. As to the 24 executives, there was only one statistically significant relation ($\chi^2 < .05$). At the group level, the correlation between confidence in decisions and

receiving or expecting to receive feedback is 0.07 (ns). In addition, levels of confidence in decisions were both unrelated to and did not vary as a function of whether feedback was received shortly after making decisions ($r = 0.08$, ns), or much later ($r = -0.07$, ns).

Attempts to find relations in the data that would “explain” confidence in decisions taken proved unsuccessful. On the other hand, there is a strong positive relation between the confidence participants have that their decisions are “right” and the confidence they express in the appropriateness of the feedback they receive or expect to receive ($r = 0.77$, $p < .01$). This therefore raises the possibility of “explaining” confidence in feedback. To what extent could this be stimulated by prior feelings of confidence in decisions taken and other variables? To explore this issue, I regressed confidence-in-feedback on confidence-in-decisions as well as other variables. As well as a positive significant effect for confidence-in-decisions ($t = 7.59$), this analysis revealed a significant effect for frequency of decisions, that is, their “habitual” nature ($t = 2.23$). The R^2 from this regression equation was 0.65, $F = 29.1$, $p < .001$. Whereas extreme care should be exercised in interpreting such a regression, it does suggest a link between confidence and the frequency of certain kinds of decisions; in other words, the more habitual, the more confident.

Discussion of results

The study illuminates four different but related issues. The first concerns differences between the decision environments of executives and students; the second is whether people receive feedback in respect of their decisions, and if so, how much; the third is expectation of feedback on decisions taken; and the fourth is the general topic of confidence in decision making. A fifth issue involves differences between data collected concurrently (by the ESM) and retrospectively (the student controls).

The decision environment of the executives and students varied significantly in both content and structure. In terms of content, students perceived the majority of their decisions to be private in nature (69%) whereas executives classified most of their decisions as professional (61%).¹¹ Second, even though about one third of the decisions of both executives and students concerned their “basic occupation,” executives had a further 19% in the area of professional communication as against 2% for the students.¹² In addition, about one quarter (27%) of the executives’ activities when they received the SMS messages occurred while they were engaged in some form of professional communication. Indeed, communication has often been quoted as a key managerial skill (see, e.g., Goleman, 1998) and one that has to be developed in younger executives. To the extent that these data are representative, it suggests an important gap between demands in the decision making environments of executives and students and thus possibly why training in communication is so important for younger executives.

As to the structure of decisions, executives clearly saw these as more complex as evidenced by the fact that whereas 80% of students’ decisions were perceived to have a positive focus, the comparable figure for executives was 58%. Students and executives also varied on reported frequency of the kinds of decisions they took. For students, 60% of decisions were described as occurring frequently (whether they “really” thought about them or not), whereas for executives the comparable figure was 42% (see Table 22.2). Both executives and students received or expected to receive feedback for about 60% of their decisions. However, for executives, feedback was more delayed in time than it was for students.

An anonymous referee suggested that one possible explanation for differences between the inferred decision environments of executives and students

could be a reporting bias, that is, the two groups differed in what they considered appropriate “DA’s” (decision actions) to report. In particular, it was argued, executives might be reluctant to report “trivial” decisions (such as taking a coffee). Whereas it is hard to reject this explanation definitively, a reading of the ESM questionnaire responses does not suggest that the executives refrained from reporting “trivial” decisions. Indeed, these form a large part of their data.

As noted above, some 40% of decisions involved no actual or expected feedback. Whether this figure is high or low is unclear because, to the best of my knowledge, no other studies have attempted to make such an estimate on the basis of random samples of people’s behavior. In addition, evidence from tacit learning would suggest that people may not always be aware of feedback and its effects (see, e.g., Reber, 1993). However, the significance of the 40% estimate is that it emphasizes the informational poverty of environments in which we learn about our decision making effectiveness. In many cases, I suspect, people simply rely on internal feelings to assess whether their decisions are correct – feelings that are probably based on having received no negative feedback in similar situations in the past. However, it should be clear that this strategy is liable to lead to self-fulfilling beliefs and actions. This, in turn, raises the important issue of seeking means to improve the level of feedback that can be obtained following actions and of making people aware of this necessity (Hogarth, 2001).

What kind of feedback do people expect to test the validity of their actions? Is this only confirmatory in nature? First, note that for decisions for which feedback has already been received (the “current” category in Table 22.4), the vast majority of actions were accompanied by confirmatory evidence. However, given the manner in

which the study was conducted, this feedback must have been received shortly after the actions were taken.

As to expected feedback, students and executives do expect confirmation (cell *a*) for 22% and 15% of their decisions, respectively. However, they indicate significant awareness that their actions could also result in negative or disconfirmatory feedback. Indeed, more than half (56%) of all the executive feedback falls in the expected *a* and *b* category. (The figure for students is 36%.) There is also occasional, albeit minimal reference to potential evidence from cells *c* and *d*. Curiously, some 9% of the feedback executives claimed they would receive was “unobservable.”

The overall picture that emerges from these data does not match a stereotype where people *only* think of confirming evidence. Both executives and students are clearly aware that not all actions will result in positive feedback. However, participants don't always indicate that they know what feedback is appropriate to assessing the validity of the actions they have taken.

Of course, in the present work there are no data that can test whether the levels of confidence exhibited by the participants are justified. A priori, the lack of a direct relation between feedback and confidence is some cause for concern. Do participants really discriminate between situations where they are or are not “justified” in expressing confidence? As speculated by Einhorn and Hogarth (1978), the mere fact of making judgments and decisions without receiving direct disconfirming evidence may be sufficient to both create and sustain feelings of confidence. In addition, illusions of confidence may sometimes have beneficial effects in that they encourage taking positive actions across time (cf., Taylor & Brown, 1988; 1994). A critical issue raised, therefore, by this research is to determine the bases on

which people establish feelings of confidence in the decisions they take. Two important sources of data not considered here are, first, people's individual histories for certain types of decisions,¹³ and second, possible individual differences in general levels of confidence.

The task given to the students who completed the retrospective questionnaire was not easy and thus the fact that responses were different from the concurrent ESM data should not be surprising. The most interesting differences concerned confidence and feedback. In the retrospective study, students reported being more confident in their decisions than their ESM counterparts, receiving more feedback, and, on average, receiving feedback with greater delays. Whereas above I interpreted the difference in confidence as contradicting research in the calibration paradigm that would lead to expecting the opposite finding (i.e., that retrospective would be less confident than concurrent, cf. Gigerenzer, Hoffrage, & Kleinbölting, 1991), I caution against emphasizing this result. First, it should be replicated in a study where both retrospective and concurrent data are collected from the same participants. Second, the real interest in the difference between concurrent and retrospective data in the present research is to demonstrate that they do not lead to the same results (cf. Ericsson & Simon, 1984; McFarlane, Martin, & Williams, 1988).

General discussion

To the best of my knowledge, no other studies have used the ESM methodology for examining decision behavior. Thus, there are no benchmarks for considering the results reported in this study. For example, should we be surprised that participants receive or expect to receive feedback on about 60% of their decisions, or that they are confident about 70%? What do these estimates imply in terms of how the people acquire decision making skills? Given that this is the first

ESM study on decision making, I limit myself here to its promise for studying further aspects of decision behavior. I first consider some methodological issues, and then turn to substantive ones.

In terms of methodology, the present study raises a number of concerns. The major concern can be summarized under the heading of selection biases. It was clear that not all executive participants were able to complete the task in the manner requested. This therefore raises the possibility that only executives with certain kinds of decision environments could complete the demands of the study and the question of how to overcome this problem in future studies. Jobs undoubtedly differ in the extent to which they allow executives to be more or less available to respond to questionnaires when they receive SMS messages. In addition, the effort needed to respond to each message can be a barrier. Some suggestions to overcome these difficulties include reducing the number of days over which studies are conducted and/or the number of messages sent per day. Simplification of the questionnaires is also a possibility. For example, questions could possibly be designed in more of a check-list format. In addition, it may be feasible to enlist technology to help facilitate the process. For example, a couple of executives in the present study transformed the questionnaire into a spreadsheet and used this to record responses. Specially programmed pocket computers have been used to collect data in other ESM studies (see, e.g., Teuchmann, Totterdell, & Parker, 1999). The use of such technology also increases the feasibility of being able to ask people different questions on different occasions and/or being able to tailor questions to prior responses on a real-time basis.

A second form of selection bias centers on which particular decisions participants chose to report. In this study, participants were requested to focus on the *decision action* that was closest in time to the moment the message was received.

However, this instruction still left much leeway to the participants to select or avoid specific types of decisions. For example, although both groups reported that some 16% to 23% of their decisions were taken “without really thinking,” my guess is that the real percentage could be much higher. One way of assessing the seriousness of this bias could be to have participants “shadowed” by an investigator as in Brunswik’s (1944) study. It would be the investigator’s task to “audit” the decisions reported by the participant. Although expensive, this could possibly be done on a small sample basis.

It would, of course, be naïve to believe that the ESM is *the* solution to methodological problems in research on decision making. No methodology owns the truth. Clearly, the ESM has great potential but this potential is most likely to be realized when it is used in conjunction with other approaches. For example, ESM could be used as an adjunct measurement tool in studies of so-called “naturalistic” decision making (cf., Lipshitz, Klein, Orasanu & Salas, 2001). It is also easy to imagine it as a supplement to questionnaires such as used in studies of consumer behavior or social psychology. Its greatest value, however, would seem to be its ability to calibrate the frequency of behaviors that have been identified by other means and to assess the kinds of samples of situations that people experience in their natural environments. We all know, for example, that we do not receive feedback on all of our decisions. However, how often does this occur? How does this vary according to different conditions, i.e., by persons and/or environments? And how important is this?

In terms of substance, what this study has achieved, *inter alia*, is a glimpse into the decision environments of groups of executives and students. It will therefore be important to replicate the present study with different populations and to develop

specific hypotheses about the types of decisions they encounter. For example, it would be intriguing to investigate differences between people with varying levels of experience in specific occupations or people with different functional responsibilities.

As to future studies, several questions would seem well-suited to the ESM. One is the relation between mood or emotions and types of decision taken. To what extent do people use different ways of making decisions when they are in “good” or “bad” moods? Mood has often been a dependent variable in ESM work and can be measured quite easily (see, e.g., Csikszentmihalyi, 1990). However, is it related to confidence in decisions made or the kind of information that people might expect as feedback? Can one detect a relation between mood and risk-taking in people’s naturally occurring environments (cf. Isen, 1993)? A further topic concerns how people make decisions, e.g., analytically, using specific heuristics, through feelings and intuitions, or some combination of the preceding. Assuming that one could teach people how to classify decisions just taken, it would be intriguing to learn more about the frequencies of different “methods” as well as their correlates, e.g., confidence, perceived risk, and so on. One difficulty, of course, with the present approach is that it is difficult to assess whether people are or are not taking “good” decisions. Finding ways to assess this represents a daunting challenge for future research.

A criticism of the present study could be that most of the approximately 1,200 decisions examined had trivial consequences and could thus be dismissed as irrelevant. It is important to recognize, however, that life’s consequences do not depend solely on how people make important decisions but may be more significantly affected by the *cumulative* effects of small and seemingly irrelevant decisions. There are two reasons. First, the sheer size of the cumulative consequences of trivial decisions can be huge. Second, even though people may use somewhat different

processes when making important as opposed to trivial decisions, the former are undoubtedly affected by the latter. It is difficult to suspend habits or routines that have built up over years of experience. Indeed, a case could be made that even incremental improvements in the manner in which people make small, everyday decisions could have huge effects on the outcomes they experience in life (see also Hogarth, 2001).

In conclusion, the present study demonstrates the feasibility of using the ESM to study decision behavior. Clearly, it cannot examine all issues that have been studied by other research methods. However, it can illuminate the kinds of decision environments that people really experience and thereby clarify possible misconceptions of what is involved in ongoing, everyday decision making. In other words, what are the samples of tasks that constrain responses of “intuitive statisticians”? Moreover, the ESM can help researchers sample decision making behaviors in ways that can illuminate what is and what is not important. For example, it has been argued that judgmental “biases” can be the result of “heuristic” strategies that are generally useful (cf. Tversky & Kahneman, 1974). However, how often do people encounter situations in which heuristics are dysfunctional? As this chapter demonstrates, we have the means to sample people, behavior, and circumstances and thus to answer these kinds of questions by achieving a more accurate understanding of the natural ecology of decision making.

Footnotes

* This research was financed partially by grants from the Spanish Ministerio de Ciencia y Tecnología, CREI at Universitat Pompeu Fabra, and CEDEP (Fontainebleau, France).

* The author is ICREA Research Professor at Universitat Pompeu Fabra, Barcelona. He is most grateful to Irina Cojuharenco and Carlos Trujillo for their excellent research assistance and to illuminating comments from Antonio Cabrales, Irina Cojuharenco, Klaus Fiedler, Peter Juslin, Jochen Reb, and an anonymous reviewer. Contact address: robin.hogarth@upf.edu

¹ See also the discussion in Goldstein et al. (2001), pp. 186-187.

² By “size constancy” is meant the ability of the perceptual system to see objects as having approximately constant size despite the fact that their projection onto the retina varies as a function of distance and other conditions.

³ I am grateful to an anonymous referee for suggesting that this study be conducted to complement results from the ESM.

⁴ Both groups of executives received feedback in the form of a report summarizing results of the study. The author also volunteered to hold a meeting at which the results could be discussed. At the time of writing, this had not yet occurred.

⁵ Copies of the instructions and the questionnaires as well as the coding scheme for qualitative data (see below) may be obtained from the author.

⁶ The second group of students is, of course, not a control group in the accepted use of this term (i.e., allocation of participants to the first and second groups was not made at random). What the second group does represent is simply a number of students drawn from the same population as those who participated in the ESM study. In a future

study, it would be of interest to replicate the present experiments by allocating respondents at random to the two groups.

⁷ This does not, of course, have to be the case because participants (and particularly the executives) varied in the number of responses they provided.

⁸ Here, as in other cases where qualitative responses were provided, I report the initial inter-coder agreement rate on classification (i.e., prior to the reaching of a consensus).

⁹ All comparisons between the distributions of responses of the student controls and the data collected concurrently from the students in the ESM study are significantly different (using χ^2 tests, $p < .01$).

¹⁰ This “outlier” was a 28-year old graduate student and quite different from the undergraduate population. Incidentally, statistical tests of any differences between the distributions of responses of students and executives on this question revealed no significant differences irrespective of whether the outlier was included.

¹¹ An anonymous referee correctly noted that comparisons between executives and students might be biased because the times for which the two groups were sampled were not identical (9 am to 7 pm, Mondays through Fridays, for the executives; and 9 am to 9 pm, Mondays through Saturdays, for the students). To assess this bias, the data from all messages sent to students after 7 pm and on Saturdays were eliminated. Across all questions, the differences between the original data (i.e., all responses) and responses left after eliminating the non-comparable subset were minimal. The only differences that might merit attention were that the subset of data indicated that more time was spent on “basic occupation” (46% vs. 41%) and, correspondingly, more decisions were professional in nature (31% vs. 26%). If anything, these results are the opposite of what one might expect (presumably the subset contained less leisure time for the students). To conclude, there might well be important differences between how

the executives and students responded to the SMS messages. However, such differences can not be attributed to the different times during which responses were elicited.

¹² Note too, that even if one sums the categories of professional and personal communication, the executives both spend more time involved in this activity than the students (31.7% vs. 6.7%) and take relatively more decisions (23.5% vs. 11.6%).

¹³ It was, of course, possible to categorize the data collected by domains of applications and comparative frequencies of decisions (see Tables 22.1 and 22.2). However, neither of these categories is sufficient to characterize a person's decision making history for kinds of decisions.

References

- Brunswik, E. (1944). Distal focusing of perception: Size constancy in a representative sample of situations. *Psychological Monographs*, 56 (254), 1-49.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.
- Camerer, C. F. (1981). *The validity and utility of expert judgment*. PhD dissertation, The University of Chicago.
- Chapanis, A. (1961). Men, machines, and models. *American Psychologist*, 16, 113-131.
- Chapanis, A. (1967). The relevance of laboratory studies to practical situations. *Ergonomics*, 10, 557-577.
- Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30, 116-127.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper-Collins.
- Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease*, 175, 526-536.
- Ebbesen, E. B., & Konečni, V. J. (1980). On the external validity of decision-making research: What do we know about decisions in the real world? In T. S. Wallsten (Ed.), *Cognitive processes in choice and decision behavior*. Hillsdale, NJ: Erlbaum.
- Einhorn, H. J. (1980). Learning from experience and suboptimal rules in decision

- making. In T. S. Wallsten (Ed.), *Cognitive processes in choice and decision behavior*. Hillsdale, NJ: Erlbaum.
- Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, 85, 395-416.
- Ericsson, A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, 107, 659-676.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98, 506-528.
- Goldstein, D. G., et al. (2001). Group report: Why and when do simple heuristics work? In G. Gigerenzer & R. Selten (eds.), *Bounded rationality: The adaptive toolbox*. Cambridge, MA: The MIT Press.
- Goleman, D. (1998). *Working with emotional intelligence*. New York: Bantam.
- Hammond, K. R. (1978). *Psychology's scientific revolution: Is it in danger?* Reprint 211, Center for Research on Judgment and Policy, University of Colorado, Boulder, CO.
- Harvey, N. (1997). Confidence in judgment. *Trends in Cognitive Science*, 1, 78-82.
- Hogarth, R. M. (1986). Generalization in decision research: The role of formal models. *IEEE Transactions in Systems, Man, and Cybernetics*, SMC-16 (3), 439-449.
- Hogarth, R. M. (2001). *Educating intuition*. Chicago, IL: The University of Chicago Press.

- Hurlburt, R. T. (1997). Randomly sampling thinking in the natural environment. *Journal of Consulting and Clinical Psychology*, 67 (6), 941-949.
- Isen, A. (1993). Positive affect and decision making. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 261-277). New York: Guilford.
- Juslin, P. (1993). An explanation of the hard-easy effect in studies of realism of confidence in one's general knowledge. *European Journal of Cognitive Psychology*, 5, 55-71.
- Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes*, 57 (2), 226-246.
- Juslin, P., Winman, A., & Olsson, H. (2000). Naïve empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect. *Psychological Review*, 107, 384-396.
- Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94, 211-228.
- Klayman, J., Soll, J. B., González-Vallejo, C., & Barlas, S. Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes*, 79 (3), 216-247.
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp.306-334). Cambridge, England: Cambridge University Press.
- Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making*, 14 (5), 331-352.

- McFarlane, J., Martin, C. L., & Williams, T. M. (1988). Mood fluctuations: Women versus men and menstrual versus other cycles. *Psychology of Women Quarterly*, 12, 201-223.
- Murphy, A., & Winkler, R. L. (1984). Probability forecasting in meteorology. *Journal of the American Statistical Association*, 79, 489-500.
- Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.
- Russo, J. E., & Schoemaker, P. J. H. (2002). *Winning decisions*. New York: Doubleday.
- Schwartz, B. (1982). Reinforcement-induced behavioral stereotypy: How not to teach people to discover rules. *Journal of Experimental Psychology: General*, 111, 23-59.
- Sjöberg, L., & Magneberg, R. (1990). Action and emotion in everyday life. *Scandinavian Journal of Psychology*, 31, 9-27.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103 (2), 193-210.
- Taylor, S. E., & Brown, J. D. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological Bulletin*, 116 (1), 21-27.
- Teuchmann, K., Totterdell, P., & Parker, S. K. (1999). Rushed, unhappy, and drained: An experience sampling study of relations between time pressure, perceived control, mood, and emotional exhaustion in a group of accountants. *Journal of Occupational Health Psychology*, 4 (1), 37-54.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 211, 453-458.

Table 22.1 -- Current activities and domains of decisions

	<u>Executives</u>			
	<u>Domain of decisions*</u>	<u>Current activity</u>	<u>Domain of decisions</u>	<u>Current activity</u>
			<u>%</u>	<u>%</u>
1. Basic occupation ***	217	208	36	34
2. Professional communication	117	169	19	27
3. Eating and drinking	61	52	10	8
4. Housework, personal time and funds management	93	49	15	8
5. Transportation	25	48	4	8
6. Personal communication	27	27	4	4
7. Acquiring information	11	22	2	4
8. Entertainment and sports	14	16	2	3
9. Sleep, rest, recreation	20	15	3	2
10. Developing additional skills	7	11	1	2
11. Personal hygiene, beautification, dressing	13	2	2	0
12. Ethics	<u>6</u>	<u>0</u>	1	0
	<u>611</u>	<u>619</u>		

	<u>Students</u>				
	<u>Domain of Decisions**</u>	<u>Current activity</u>	<u>Domain of decisions</u>	<u>Current activity</u>	<u>Activities of controls</u>
			<u>%</u>	<u>%</u>	<u>%</u>
1. Basic occupation ***	173	243	30	41	30
2. Eating and drinking	107	95	18	16	12
3. Sleep, rest, recreation	50	56	9	10	14
4. Housework, personal time and funds management	61	39	10	7	4
5. Transportation	26	38	4	6	7
6. Personal communication	60	36	10	6	8
7. Entertainment and sports	26	22	4	4	7
8. Developing additional skills	19	32	3	5	3
9. Personal hygiene, beautification, dressing	29	10	5	2	4
10. Acquiring information	17	11	3	2	6
11. Professional communication	13	6	2	1	2
12. Ethics	<u>1</u>	<u>0</u>	0	0	3
	<u>582</u>	<u>588</u>			

* inter-coder agreement: 76%.

** inter-coder agreement: 81%.

*** the data are ordered by frequency of current activity

Table 22.2 -- Characteristics of decisions

	<u>Executives</u>	<u>Students</u>	<u>Total</u>	<u>Executives</u>	<u>Students</u>	<u>Student controls</u>
<u>Action type</u>				<u>%</u>	<u>%</u>	<u>%</u>
Positive focus	358	453	811	58	80	x
Negative focus	36	27	63	6	5	x
Multiple alternatives	<u>218</u>	<u>89</u>	<u>307</u>	36	16	x
	<u>612*</u>	<u>569**</u>	<u>1.181</u>			
 <u>Orientation</u>						
Professional	375	153	528	61	26	28
Private	222	397	619	36	69	56
Both	<u>15</u>	<u>28</u>	<u>43</u>	2	5	16
	<u>612</u>	<u>578</u>	<u>1.190</u>			
 <u>Frequency: Was this something that you do</u>						
Frequently and without really thinking?	99	134	233	16	23	34
Frequently but you do think about it?	161	213	374	26	37	45
From time to time?	197	122	319	32	21	12
Rarely?	96	80	176	16	14	6
This was the first time!	<u>58</u>	<u>34</u>	<u>92</u>	9	6	3
	<u>611</u>	<u>583</u>	<u>1.194</u>			
 <u>Confidence in the "right" decision</u>						
Very confident	189	203	392	31	35	42
Confident	232	181	413	38	31	32
Somewhat confident	150	77	227	25	13	16
Not confident	35	103	138	6	18	7
Not at all confident	<u>6</u>	<u>20</u>	<u>26</u>	1	3	3
	<u>612</u>	<u>584</u>	<u>1.196</u>			

* inter-coder agreement: 90%.

** inter-coder agreement: 93%.

Table 22.3 -- Feedback: time and confidence

	<u>Executives</u>	<u>Students</u>	<u>Total</u>	<u>Executives</u> %	<u>Students</u> %	<u>Student</u> <u>controls</u> %
<u>Feedback</u>						
Yes	395	305	700	65	52	70
No	<u>217</u>	<u>279</u>	<u>496</u>	35	48	30
	<u>612</u>	<u>584</u>	<u>1,196</u>			
<u>Time until feedback</u>						
Less than 2 minutes	80	96	176	20	32	17
2 to 5 minutes	3	3	6	1	1	8
5 to 15 minutes	16	31	47	4	10	4
15 to 45 minutes	17	35	52	4	12	11
45 to 90 minutes	18	13	31	5	4	8
90 minutes to 3 hours	8	21	29	2	7	7
3 hours to 15 hours	24	21	45	6	7	5
15 hours to 2 days	21	12	33	5	4	15
2 days to 1 week	30	9	39	8	3	7
1 week to 1 month	56	21	77	14	7	8
More than 1 month	81	40	121	21	13	11
Indeterminate	<u>41</u>	<u>2</u>	<u>43</u>	10	1	0
	<u>395</u>	<u>304</u>	<u>699</u>			
<u>Confidence in feedback</u>						
Very confident	179	119	298	46	40	35
Confident	130	102	232	33	34	32
Somewhat confident	77	26	103	20	9	23
Not confident	6	43	49	2	15	7
Not at all confident	<u>0</u>	<u>6</u>	<u>6</u>	0	2	3
	<u>392</u>	<u>296</u>	<u>688</u>			

Table 22. 4 -- Type of feedback

<u>Classification</u>		<u>Executives*</u>	<u>Students**</u>	<u>Total</u>	<u>Executives</u>	<u>Students</u>	<u>Student</u>
					<u>%</u>	<u>%</u>	<u>controls***</u>
					<u>%</u>	<u>%</u>	<u>%</u>
Unobservable		36	9	45	9	3	0
Current	Cell a	60	88	148	14	29	22
	Cell b	1	8	9	0	3	9
	Cell c	1	0	1	0	0	0
	Cell d	4	1	5	1	0	0
Expected	Cell a	61	67	128	15	22	35
	Cell b	9	12	21	2	4	17
	Cell c	2	1	3	0	0	1
	Cell d	4	4	8	1	1	1
	Cell a or b	233	110	343	56	36	12
	Cell c or d	<u>7</u>	<u>6</u>	<u>13</u>	2	2	1
		<u>418</u>	<u>306</u>	<u>724</u>			

* inter-coder agreement: 78%

** inter-coder agreement: 79%

*** responses normalized to facilitate comparisons

Figure caption

Figure 1 – Types of feedback

	<u>Feedback</u>	
	<u>Confirming</u>	<u>Disconfirming</u>
<u>Action taken</u>	Cell a	Cell b
<u>Action not taken</u> *	Cell c	Cell d

* This could mean that another action was taken as opposed to no action having been taken.