ISSUES AND POLICY

The consistency of UK pension fund trustee decision-making

GORDON L. CLARK*, EMIKO CAERLEWY-SMITH, and JOHN C. MARSHALL†

Oxford University Centre for the Environment, South Parks Rd., Oxford OX1 3QY, UK and *Labor and Worklife Program, Harvard Law School, Harvard University, Cambridge MA 02138, USA; †Department of Clinical Neurology, Oxford University, Radcliffe Infirmary, Oxford OX2 6HE, UK.

Abstract

An important research programme in the social sciences concerns the theory and practice of individual decision-making under conditions of risk and uncertainty. At the same time, it is apparent that western governments increasingly rely upon individuals to plan and maintain savings programmes to meet their income aspirations: the long-term retirement income of those outside of the welfare state depends a great deal on the competence and consistency of individual decision-making. In this paper, we use a set of problems requiring the same techniques of judgement to test the consistency of trustee decision-making. Respondents were a group of trustees drawn from select UK defined benefit pension plans compared with a larger group of Oxford undergraduates. It was found that many respondents were inconsistent across related problems requiring the application of probabilistic judgement. It is also shown that trustees were more consistent than many undergraduates and it appears that trustee education and professional qualifications can make a positive difference to consistent decision-making. A more challenging test that depends upon understanding the relationship between demographic ageing, immigration, and the financing of pay-as-you-go social security suggests that substantive knowledge and consistency of judgement are crucial components of expertise. Implications are drawn for the trustee institution and the wider debate over the role and significance of individual decision-making with respect to income aspirations.

Introduction

One of the most important research programmes in the social sciences concerns the nature and scope of human rationality; this research programme is largely
empirical and challenges conventional assumptions about the rational actor model. Using the methods and tests of psychology, Thaler (1985) amongst others has shown that many people act in ways inconsistent with the axioms of expected utility maximization models. So, for example, it has been shown that people are surprisingly risk-averse, preferring the certainty of a low return as opposed to the uncertain prospect of a much higher return. This research programme has important implications for the social sciences and especially for economics and business studies where strong conceptions of agent rationality joined with the efficient markets hypothesis have tended to exclude consideration of ‘the norms, habits, and personal characteristics of the decision-maker’ (Tversky and Kahneman, 1981: 453).

Whatever the significance of rationality for economics, other social sciences are less committed to this frame of reference, especially when considering the interaction between human behaviour and culture (anthropology), context (geography), and society (sociology) (see Henrich et al., 2005). By contrast, much research in psychology and economics has been focused on supposed universals of human cognition, including neurological studies of how the brain works in response to identifiable stimuli and tasks (for a recent survey see Camerer et al., 2005). Some years ago, however, Herbert Simon (1956) noted that cognition is but one blade of a pair of scissors; the other blade is the environment, implying that human behaviour is to be found at the intersection of both blades where cutting paper, that is, dealing with real life problems, requires judgement. This is a remarkable metaphor even if it has proven difficult to extend studies of cognition to individual roles and responsibilities (but see Bröder, 2003).

Elsewhere we have sketched an approach to rationality grounded in neuropsychology (Clark and Marshall, 2002), introduced issues related to the governance of pension fund decision-making (Clark, 2004), and reported the results of tests of the competence of pension fund trustees with respect to investment-related problems (Clark et al., 2006). In doing so, recognized tests of decision-making have been integrated with the trust institution, making an explicit connection between tests of individual competence and institutional roles and responsibilities. As a point of reference, these tests of competence were administered to Oxford undergraduates who share neither the age and experience of trustees nor their roles and responsibilities. Nonetheless, it was found that trustees and undergraduates share the same kinds of ‘biases’ or ‘habits of mind’ identified in other similar studies about the cognitive foundations of decision-making (see Lo, 2004). It was also found that the range of trustee competence was far wider than expected, challenging assumptions made about the coherence of pension fund decision-making.

Consistency is a crucial element of the theory of utility maximization; people should evaluate options in a consistent manner, taking into account the value and probability of expected outcomes. But, in many cases, people can neither quantify the expected value of anticipated outcomes nor apply realistic probabilities to likely outcomes because information is of poor quality, expensive to obtain, or simply unavailable. The issue is the consistency of judgement type (or procedure) rather than consistency of results (or substance), a distinction exploited by March (1994) and Simon (1982) amongst others. This paper is about the extent to which respondents
are consistent in their judgement across the same kinds of problems formulated in different ways. The responses of trustees are compared with those of Oxford undergraduates, seeking evidence for the effects that education, professional qualifications, and task-specific training have on the degree of consistency of judgement.\(^1\) We also test for consistency of judgement with respect to the relationship between demographic ageing, immigration, and the financing of pay-as-you-go social security.\(^2\)

As in previous work, it is shown that trustees differ in terms of their use of probabilistic reasoning across common problems. Furthermore, they do not agree on the relationship between demographic ageing, immigration, and the financing of state pensions. On the other hand, they are more consistent on these issues than Oxford undergraduates, especially when we take into account trustee education and professional qualifications. These results have implications for the selection of trustees. They also have implications for how trustees might integrate decision-techniques (for example, the application of probabilistic reasoning) with an appreciation of the causal relationships relating to the long-term financing of UK pensions. It is contended that there is a difference between competence (solving problems) and expertise (integrating decision-technique with causal analysis); only a minority of trustees could be said to be experts. These results have implications for the governance of pension funds. But they also have implications for individual responsibility in planning their retirement wellbeing.

In the next section, the nature and scope of rationality is considered before turning in the subsequent section to the methodology used to test consistency. Thereafter the results of analysis are presented, focusing upon consistency of technique and then consistency with respect to causal relationships. The paper is brought to a close with a section devoted to implications and conclusions. The reach and significance of our conclusions should be qualified on two counts. First, as is discussed below, the sample of trustees used in the analysis is small and drawn from a select group of pension funds. Second, the testing regime is more ambitious than most, making tentative links between respondents’ cognitive abilities, their roles and responsibilities, and their personal characteristics. Our results are suggestive of further research rather than of definitive conclusions about the consistency of individual decision-making in a specific domain.

**Rationality and decision-making**

Basic rationality – logical deduction from first principles – is an innate human trait. It could hardly be otherwise given acknowledged instincts of self-preservation

---

\(^1\) Here, we refer to measures of educational attainment and recognized professional qualifications. These types of personal ‘qualities’ are widely known. Perhaps less appreciated but no less important is the fact that generic skills such as communication, technical know-how, literacy, and numeracy are correlated with educational attainment and professional qualifications (see the Interim Report of the UK Leitch Review of Skills 2005).

\(^2\) Throughout the paper, our focus is upon the empirical dimensions of the consistency of decision-making. A most important issue is the consistency of decision-making as a matter of social justice: applying rules and procedures in a manner that is consistent across alike and non-alike cases. This issue is most evident, at present, in debate over the validity of immigration appeals processes in a number of countries, including Canada, Australia, and the United Kingdom.
including the protection of children as well as the vast body of research that has proven time and again that people respond to incentives and sanctions in all kinds of situations. However, complex logical reasoning as a cognitive ability is neither equally shared amongst human beings nor necessarily determinate of behaviour in specific circumstances. Just as some people can jump higher, run further, and excel at feats of endurance, evidence suggests some people are just better than others at logical reasoning. Even so, sheer biology is hardly sufficient as an explanation of individual behaviour and relative performance: we must also factor in issues such as individuals’ skill-sets and preferences for various socially defined measures of esteem. These issues are, arguably, just as important as biological imperatives.

Economist/psychologists such as March (1994) and Simon (1982) have argued that human behaviour is the product of cognitive ability and the environment – given shape by Simon’s scissors metaphor which suggests interaction and precision. Other analysts such as Tversky and Kahneman (1981) have suggested that agents carry with them certain predispositions regarding risk-taking and that the context of decision-making is an essential ingredient in the process of choosing amongst alternative courses of action. In their work, ‘framing’ is the metaphor and could be thought to stand for Simon’s ‘environment’. What is striking about the research programme underpinning these arguments is the fact that it is based upon respondents’ solutions to well-defined problems tested many times over. It is an empirical regime designed to calibrate the nature and scope of rationality; it is neither idealistic about decision-making nor definitive about what behaviour could be or ought to be.

For social scientists committed to an empirically founded conception of rationality, the focus of research is increasingly upon the methods and techniques of decision-making. By implication, the degree to which economic agents optimize with respect to desired outcomes is an empirical matter rather than an a priori assumption. This research programme makes three assumptions about the status of the environment (culture, context, and society) in decision-making. First, decision-making takes place under risk and uncertainty such that judgements are made amongst various probable outcomes (risk) and with recognition that there are unspecified events that may disrupt and change current circumstances (uncertainty). Second, strategic thinking is an important attribute of effective decision-making; being able to revise one’s goals and switch from planned actions to other unanticipated courses of action could be just as important as long-term planning. Third, consistency in the application of effective techniques of decision-making rather than consistency of decisions with respect to given preferences may be the mark of expert decision-making.

Few people, it seems, are self-conscious decision-makers; many rely on inchoate rules of thumb or templates to prompt action in circumstances demanding response. A century ago, Sidgwick (1907 [1981]) suggested that intuition was the governing

3 Debate about the ontological status of rationality is found mostly in philosophy. See, for example, Davidson (2004: 196–197) disputing the very possibility of people not being rational. Using logic, he argues that being and deciding is sufficient to suggest that people are rational and that ‘[a]n agent cannot fail to comport most of the time with the basic norms of rationality’. This is a large claim, considering the lack of any evidence to demonstrate how and why this is true.
behavioural principle in these situations – intuition being an automatic response to circumstances based upon past experience filtered through some kind of elementary coding device that determines the relevance of one response over another.\(^4\) As described, intuition has an information-processing function like a modern-day cognitive model. We make no such assumption of equivalence. Rather, we simply emphasize that most people bring to each and every decision-situation a tool bag of techniques whose relevance is determined in the light of experience (Gigerenzer et al., 1999). It is also the case, however, that specific kinds of decision-environments may demand certain kinds of decision-techniques; being able to recognize the nature of the problem is a necessary step in choosing the appropriate decision-technique.

In this case, consistent decision-making requires a modicum of comprehension. At one level, comprehension could be just an issue of pattern recognition and response. Comprehension could also be the product of substantive knowledge calibrated against past successes and failures using decision-techniques; knowing a great deal about the underlying causal structure of a decision-environment combined with experience using relevant decision-techniques may yield solutions that are more often ‘correct’ rather than ‘incorrect’. Therefore, there are two kinds of consistency in decision-making: that which is competent (based upon pattern recognition and linked response) and that which is expert (based upon substantive knowledge and deliberate decision-technique). This carries three implications: first, even if a recognition and response technique is successful in the short-term, it may well fail in the long term without substantive knowledge; second, expertise is domain specific (Wagner, 2002); and, third, task-specific education, professional qualifications, and training may help distinguish competence from expertise (Gigerenzer and Hug, 1992).

In this paper, we analyze the consistency of decision-making, focusing on the application of decision-techniques including probability-assessment, to the same kind of problem expressed in different ways. Probability judgements have special significance in our research because it is an important decision-technique for pension fund investment decision-making and for the application of modern portfolio theory (Ho and Lee, 2004). But, in a larger sense, being able to make rudimentary probabilistic judgements is important to many people whatever their roles and responsibilities. For many issues, not just for pension and retirement income, individuals now bear the risks previously held by the nation-state (see Strauss, 2006; Venti, 2006). As in previous research, we focus on whether pension fund trustees are more or less consistent than Oxford undergraduates and the degree to which education, professional qualifications, and task-specific training make a difference to consistency.

**Experimental regime**

Social scientists tend to approach the empirical analysis of individual behaviour from one of two vantage points. Based upon third-party macro-data, behaviour is

---

\(^4\) Sidgwick’s argument and classification of the various forms of intuition have been developed and given a modern twist by Audi (2004) with reference to the philosophy of ethics. Similarly, Damasio’s (2004) argument about the role of emotions in decision-making has resonance.
inferred rather than directly observed. Sophisticated studies of panel data that allow researchers to correlate individual characteristics with observed outcomes also rely upon inference. Another approach is to survey individuals for their attitudes on certain issues and then infer behaviour from attitude. The classic example is, of course, a survey that attempts to match people’s party-political attitudes or moral beliefs with their voting behaviour. There are sophisticated ways of calibrating attitude, although it is widely recognized that inferences to behaviour can be quite problematic (Wicker, 1969).

Previously, it was shown that there is a significant difference between pension fund trustees’ attitudes to risk and the risks that they would willingly assume if calibrating the value of different kinds of pay-offs with different kinds of probabilities of success (Clark, Caerlewy-Smith, and Marshall, 2006). Many trustees reported that they would willingly assume ‘moderate’ risk, but when, in fact, asked to calibrate risk they were more often than not highly risk-averse in a manner consistent with the results of similar testing reported in the psychology literature. One explanation of the observed difference between attitude and behaviour relies upon the fact that ‘moderate’ risk-taking is the mantra of the investment management industry; by contrast, trustees are very conscious of their obligations to protect the welfare of pension beneficiaries. We also found that male undergraduates had quite liberal attitudes to risk-taking and when calibrating various risk options tended to assume more risk than trustees (a fact recognized by the insurance industry in setting insurance premiums for young male automobile drivers).

Testing regimes of problem-solving skills have a number of characteristics. First, testing regimes are often based upon well-defined and much-used problems with a considerable history in the psychology literature. Second, testing regimes tend to assess generic ability rather than skills specific to particular roles and responsibilities or for that matter appropriate to the socio-demographic characteristics of respondents. Third, tests are typically applied to undergraduate or postgraduate students; respondents who are easily enrolled and are normally responsive to puzzles with modest pay-offs (see, for example, Viscusi and Zeckhauser, 2005). Here, we begin with a testing regime that utilizes recognized problems, assessing the consistency of trustee decision-making against a set of undergraduate students (see Kovalchik et al., 2005). By focusing upon pension fund trustees, there is a risk of departing from the norms of conventional psychometric testing regimes. Nonetheless, Robertson (2001), inspired by Newell and Simon (1972) amongst others, has argued that there are good reasons to focus upon task-specific decision-techniques and the interaction between problem-solving skills and decision-environments.

This paper derives from a larger project on the competence of UK pension fund trustees sponsored by the National Association of Pension Funds (NAPF). As such, it is part of a wider debate about the governance of pension funds.5 The debate over pension fund trustee competence was initiated by Paul Myners, then the CEO of Gartmore (the UK fund manager), in a report to the Chancellor of the Exchequer. The Myners Report (2001) called for the introduction of codes of practice to enhance pension fund decision-making in a manner consistent with UK financial innovation and development. This issue was followed up in a number of government reports (see, for example, Horack et al., 2003) before the passage of the 2004 Pension Act and the introduction of regulations governing the introduction of such codes (see The
the project are available from the NAPF and have been the subject of numerous presentations to the pension fund community and the investment management industry. The recruitment of pension fund trustees to take part in the testing regime was done on the explicit understanding that their roles and responsibilities with respect to the investment of pension fund assets are significant; it was also acknowledged that these roles and responsibilities require a skill-set and level of competence so that trustees can actively engage with professional financial service providers. The Pensions Act (2004) refers to an appropriate level of trustee ‘knowledge and understanding’, while not prescribing minimum levels of education, professional qualifications, and job-specific training.

About 40 pension fund trustees were recruited, drawn from a select group of large UK defined benefit pension funds, including both public and private sector institutions. Trustees were male, typically more than 50 years of age, with higher levels of education than the average UK citizen, but with quite variable professional qualifications. Some trustees had significant accounting and actuarial skills, especially those that were employer-nominated. There were slightly more member-nominated trustees than employer-nominated trustees. As for the Oxford undergraduates, about 80 students were recruited from a required research methodology class. Their average age was 19 years, a slight majority was women, and (of course) none had the roles and responsibilities of trustees, and very few had had formal education and training in probabilistic skills related to the world of investment management. Nonetheless, they were similar to trustees in terms of their socio-demographic profiles (being Caucasian, well-educated, and from middle to upper income families).

Trustees were very much aware of the context and significance of the testing regime, whereas the undergraduates were told simply that it was a useful exercise in data gathering with lessons to be learnt about the structure and format of setting puzzles. The testing regime required more time than that normally associated with measuring people’s attitudes. In many cases, trustees put aside a day to travel to Oxford and take part in the ‘puzzle day’. For trustees, the puzzle set was comprised of two papers each of which took about an hour to complete. The format and structure of the papers are available from the NAPF report. Questions pertaining to particular decision-techniques were pseudo-randomly distributed throughout the papers so that (as much as possible) respondents were not cued on the appropriate response by prior questions of a similar type.

Pensions Regulator, 2005). The challenges facing pension fund trustees have grown over the past five years, as a combination of unanticipated forces have significantly discounted the value of pension fund assets in relation to growing expected liabilities (Clark, 2006).

See the material collected at http://www.ouce.ox.ac.uk/research/spaces/programmes/trustee.php

It should be recognized that a sample of 40 trustees is small, especially when judged against social science conventions. Given the variables considered, there are real limits on the available degrees of freedom for statistical analysis. In any event, our sample size is similar to that which government agencies have used to analyse the trustee institution. Many analysts have noted the difficulty of identifying and obtaining the cooperation of trustees for attitudinal surveys let alone the ambitious scope of this study. See, for example, Bunt et al. (1998), Horack et al. (2003), and Thomas et al. (2000). Therefore, we must be cautious of the reach of our conclusions.
Calibrating consistency of decision-making

Many tests of consistency focus upon respondents’ ability to make probability judgements, often expressed in terms of percentage likelihood – a crucial skill given trustee responsibilities when assessing alternative investment options. This point is developed in Clark et al. (2006) where reference is made to the theory and practice of investment management under risk and uncertainty. Here, we are concerned with trustees’ consistency with respect to recognizing and solving similar probability-related problems with due regard to trustees’ education, professional qualifications, and task-specific training. In the following section, we also test the significance of substantive knowledge in prompting recognition of the equivalence between two rather different specifications of the same issue (the relationship between immigration and the funding of pay-as-you-go social security).

Conjunction and probability

One way of gauging the ability of respondents to make judgements about likelihood or probability is to set a series of propositions that begin with a statement: person X has certain socio-demographic characteristics and beliefs such that the likelihood of X also being Y (a professional of some kind) is J (less than the probability of being X alone). This can be repeated with further qualifications to X, suggesting that the probability of X being Y, Z etc. is successively discounted from J. This is the ‘Linda’ conjunction puzzle made famous by Tversky and Kahneman (1983) and elaborated by Baron (1994). There is evidence that people often fail to recognize that the probability of the conjunction of two propositions must be less than the probability of either proposition on its own. This is usually taken as evidence of a failure of logical reasoning (but see Gigerenzer et al., 1999).

In this study, respondents were asked a set of three inter-related ‘Linda’ problems. Their responses were first deemed either ‘correct’ or ‘incorrect’. Respondents were then allocated to one of four categories: being consistently incorrect, correct for one instance, or two instances, or consistently correct for all three instances. In Figure 1,

![Figure 1. Consistency across ‘Linda’ conjunction problems](image-url)
the results of the respondents’ responses are summarized distinguishing between trustees and undergraduates according to the four categories. While a similar proportion of trustees and undergraduates were consistently incorrect, it is apparent that trustees were more likely to be consistently correct or nearly consistently correct (two out of three correct) than undergraduates. The most frequent response of undergraduates was one instance correct, which suggests a basic lack of understanding of the problem, leading to guess-work.

In Table 1, it is shown that those trustees who were consistently correct had at least A-level high school education with the majority reporting an undergraduate university degree. A small proportion also had an M.Sc. degree. Those consistently correct also reported the highest level of professional qualifications, including accountancy and actuarial qualifications. The level of education seemed irrelevant to respondents who were consistently incorrect. Curiously, however, those consistently incorrect and those with two out of three correct both reported low levels of professional qualifications. Task-specific professional training for trusteeship also appears to be related to consistency, including getting at least two out of three instances correct. Moreover, those consistently correct had the highest levels of training (expressed in terms of the number of training sessions from a variety of sources). These results are at odds with those of Tversky and Kahneman (1983) who reported little difference between statistically naïve and sophisticated university students in their lack of ability to recognize and solve conjunction problems. In our study, there were significant differences between (naïve) university students and (experienced) trustees on this count.

Table 1. Trustee response to Linda’s conjunction by educational attainment and professional qualifications

<table>
<thead>
<tr>
<th>Linda’s conjunction problem</th>
<th>Educational attainment</th>
<th>Professional qualifications</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High school</td>
<td>University</td>
</tr>
<tr>
<td>0 Correct</td>
<td>37.5</td>
<td>37.5</td>
</tr>
<tr>
<td>1 Correct</td>
<td>62.5</td>
<td>12.5</td>
</tr>
<tr>
<td>2 Correct</td>
<td>22.2</td>
<td>44.4</td>
</tr>
<tr>
<td>3 Correct</td>
<td>30</td>
<td>70</td>
</tr>
</tbody>
</table>

Notes: a each row shows the percentage of response by category according to trustee educational attainment or professional qualifications; b high school attainment includes both O Level/GCSE and A/AS level attainment; c university attainment includes both undergraduate and masters level attainment; d pension specific qualifications include accountancy, actuarial, other pension specific, and multiple qualifications.

Note also in our study it was found that employer-nominated trustees (ENT) had higher levels of education, professional qualifications, and training than member-nominated trustees (MNT). They were also more likely to be consistently correct (although there was no difference between the proportions of ENT and MNT consistently incorrect).
Problem recognition and probability

An elementary aspect of probability theory is that summation of the possible outcomes in a given situation cannot add to more than 1 (or 100% if expressed in terms of likelihood). For example, in any football game if Team A has a 0.6 probability of winning then Team B cannot have a greater than 0.4 probability of winning. This basic rule of probability was tested in three supposedly different situations, each requiring the same resolution (that is, the sum of $\Pr \leq 1.0$). Respondents were asked to deal with a desert camel race with four camels, to predict the distribution of MPs by political party after the (then) forthcoming UK election, and to assess the likelihood that the share price of BP would increase, decrease, or remain the same over a certain time period. Respondents were also asked to respond to the MP question using fractions instead of percentages.

Table 2 summarizes the pattern of correct responses by problem and by respondent type. Trustees did better on three of the four questions and most significantly on the BP and MP problems, whereas undergraduates did best only on the camel racing problem. The BP and the MP problems were obviously relevant to trustee roles and responsibilities. The only possibly ‘relevant’ problem for undergraduates was the MP question; the camel racing question was perhaps the most obscure but the best answered by undergraduates. Neither group did well on the MP fractions problem, although it was logically the same as the MP (%) problem. In terms of consistency, Figure 2 summarizes the results comparing the two groups of respondents according to the number of problems solved correctly. There is a remarkable difference between the two groups: trustees were far more likely to be consistently or nearly consistently correct, although the proportion of trustees consistently correct was only 30% of trustees. Again, the results suggest that undergraduates tended not to understand the basic principle.

Table 2. Percent respondents correct for sub-additivity problems by respondent type

<table>
<thead>
<tr>
<th>Sub-additivity problems</th>
<th>Respondent type</th>
<th>MPs</th>
<th>BP shares</th>
<th>Camels</th>
<th>MPs (fractions)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trustees</td>
<td>71.1</td>
<td>82.1</td>
<td>61.1</td>
<td>42.1</td>
</tr>
<tr>
<td></td>
<td>Undergraduates</td>
<td>53.4</td>
<td>58.4</td>
<td>68.1</td>
<td>12</td>
</tr>
</tbody>
</table>

An elementary aspect of probability theory is that summation of the possible outcomes in a given situation cannot add to more than 1 (or 100% if expressed in terms of likelihood). For example, in any football game if Team A has a 0.6 probability of winning then Team B cannot have a greater than 0.4 probability of winning. This basic rule of probability was tested in three supposedly different situations, each requiring the same resolution (that is, the sum of $\Pr \leq 1.0$). Respondents were asked to deal with a desert camel race with four camels, to predict the distribution of MPs by political party after the (then) forthcoming UK election, and to assess the likelihood that the share price of BP would increase, decrease, or remain the same over a certain time period. Respondents were also asked to respond to the MP question using fractions instead of percentages.

Table 2 summarizes the pattern of correct responses by problem and by respondent type. Trustees did better on three of the four questions and most significantly on the BP and MP problems, whereas undergraduates did best only on the camel racing problem. The BP and the MP problems were obviously relevant to trustee roles and responsibilities. The only possibly ‘relevant’ problem for undergraduates was the MP question; the camel racing question was perhaps the most obscure but the best answered by undergraduates. Neither group did well on the MP fractions problem, although it was logically the same as the MP (%) problem. In terms of consistency, Figure 2 summarizes the results comparing the two groups of respondents according to the number of problems solved correctly. There is a remarkable difference between the two groups: trustees were far more likely to be consistently or nearly consistently correct, although the proportion of trustees consistently correct was only 30% of trustees. Again, the results suggest that undergraduates tended not to understand the basic principle.

The degree to which trustee education, professional qualifications, and training were related to sub-additivity consistency is summarized in Table 3. Basically, trustees with limited education and no professional qualifications were likely to solve the BP problem correctly, were less likely to solve the MP and camel racing problems correctly, and were unlikely to be consistently correct. On the other hand, trustees

---

\[^{9}\] This result challenges assertions made by Gigerenzer et al. (1999) that people are better able to judge fractions rather than percentages or probabilities. Whether this is specific to these two groups of respondents or simply related to the specific nature of the problem set remains to be seen.
with university education and professional qualifications were very likely to solve the individual problems correctly and were very likely to be consistently correct across the problems. Training seemed to have a positive effect on consistency only if trustees had had three or four training sessions.

**Judgement, information, and probability**

The most challenging problems have to do with combining information to make judgements about the likelihood of events. So, for example, if we know there are 100 balls of which 50 are red and 50 are black, the base-rate Pr (black) = 0.50. If we are
Table 4. Type of reasoning used by respondents across three probability problems and the consistency of that reasoning

<table>
<thead>
<tr>
<th></th>
<th>Trustees</th>
<th>Undergraduates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number (percentage)</td>
<td>Number (percentage)</td>
</tr>
<tr>
<td></td>
<td>respondents</td>
<td>respondents</td>
</tr>
<tr>
<td><strong>Problem</strong></td>
<td><strong>Type of reasoning</strong></td>
<td><strong>Problem</strong></td>
</tr>
<tr>
<td>Jury</td>
<td>Bayesian</td>
<td>Jury</td>
</tr>
<tr>
<td>N = 39</td>
<td>6 (15.4)</td>
<td>16 (23.2)</td>
</tr>
<tr>
<td>Indicant</td>
<td>21 (53.8)</td>
<td>Indicant</td>
</tr>
<tr>
<td>Base Rate</td>
<td>12 (30.8)</td>
<td>Base Rate</td>
</tr>
<tr>
<td>Digirosa</td>
<td>Bayesian</td>
<td>Bayesian</td>
</tr>
<tr>
<td>N = 35</td>
<td>2 (5.7)</td>
<td>6 (8.5)</td>
</tr>
<tr>
<td>Indicant</td>
<td>16 (45.7)</td>
<td>Indicant</td>
</tr>
<tr>
<td>Base Rate</td>
<td>17 (48.6)</td>
<td>Base Rate</td>
</tr>
<tr>
<td>Clinical</td>
<td>Bayesian</td>
<td>Bayesian</td>
</tr>
<tr>
<td>N = 34</td>
<td>0 (0)</td>
<td>5 (7.2)</td>
</tr>
<tr>
<td>Indicant</td>
<td>24 (63.2)</td>
<td>Indicant</td>
</tr>
<tr>
<td>Base Rate</td>
<td>10 (26.3)</td>
<td>Base Rate</td>
</tr>
<tr>
<td>All</td>
<td>Consistently Bayesian</td>
<td>Consistently Bayesian</td>
</tr>
<tr>
<td>N = 39</td>
<td>0 (0)</td>
<td>1 (1.5)</td>
</tr>
<tr>
<td>Consistently Indicant</td>
<td>5 (12.8)</td>
<td>Consistently Indicant</td>
</tr>
<tr>
<td>Consistently Base Rate</td>
<td>0 (0)</td>
<td>Consistently Base Rate</td>
</tr>
<tr>
<td>Inconsistent Reasoning</td>
<td>34 (87.2)</td>
<td>Inconsistent Reasoning</td>
</tr>
</tbody>
</table>

Note: ‘Missing’ and ‘other’ responses were excluded from analysis, except in the case of ‘All’ where respondents who missed out or answered ‘other’ to any of the three questions were classified as having ‘Inconsistent Reasoning’.

In the jury problem (Table 4), 15% of trustees recognized the need to combine information to solve the problem. In the digirosa problem, just 5% of trustees recognized the Bayesian solution, while in the clinical problem none recognized the solution. By contrast, 23% of undergraduates recognized the Bayesian solution to the jury problem, 8% of undergraduates recognized the solution to the digirosa problem, and 7% of undergraduates recognized the solution to the clinical problem.

---

Baron (1994: 212–213) provides examples of related problems designed to test Bayesian inference (and more). Specifically, we use Tversky and Kahneman’s (1982) taxi problem (referred to here as the jury problem), a version of Casscell et al.’s (1978) clinical test problem, and a problem related to identifying a disease, in this case digirosa (Doherty and Mynatt, 1990).
The most common decision-technique by both sets of respondents was to use indicant information but not base-rate information (matching the results of Tversky and Kahneman and many others). When the consistency of response across the three problems was considered, it was found that no trustees and only 9% of undergraduates used either Bayesian inference or base-rate information across all three problems and just 13% of trustees and 28% of undergraduates used indicant information across all three problems.

Some analysts have doubted the appropriateness of Bayesian as opposed to indicant solutions in these types of problems (see Cohen, 1981). As well, expectations regarding the competence of decision-makers must be balanced against the habits of mind and the environment of respondents (Tversky and Kahneman, 1981: 453). From the results, it seems that for respondents to recognize the relevance of Bayesian reasoning, they would need instruction by an expert reinforced with practice (Gigerenzer, 2000: 120). In any event, no discernable relationship could be established between consistent Bayesian reasoning or otherwise with educational attainment and professional qualifications. That most respondents alert to the value of indicant information tended not to use this type of information across all problems also suggests that task-specific training may be needed to prompt recognition of the nature of these types of problems. While cautious of the sample size underpinning these conclusions, it was also found that consistent trustees had at least undergraduate education and relevant professional qualifications.

**Context, knowledge, and judgement**

To this point, it has been shown that trustee education, professional qualifications, and task-specific training are important in distinguishing between those trustees shown to be more or less consistent in their decision-making. In this respect, we followed Tversky and Kahneman and others regarding the significance of respondents’ personal characteristics relevant to their roles and responsibilities. But this type of argument can be taken further: the substantive context in which trustees make decisions is arguably an important element in trustee performance. Gigerenzer (2000), for example, suggests that the formal principles of logic and probability will always be less than adequate when explaining human behaviour (echoing claims made by Simon 1982 amongst others). The next step was to formulate a set of problems that could help determine whether context is an important ingredient in trustee decision-making.

When dealing with the likely distribution of MPs between political parties after the (then) looming UK general election, some trustees and more undergraduates violated the rule regarding summation to no more than 1 when the question was framed in terms of percentages and fractions. In fact, some experimental psychologists claim that people are better at ratios and fractions than percentages (see Tversky and Kahneman, 1981). Recognizing the significance of demographic projections, and especially estimates of the projected ratio of working to retired people for

---

11 See the essay by Gigerenzer and Selton (2001: 5–6), where they make this association explicit including reference to Simon’s scissors metaphor.
the future funding of defined benefit and state pay-as-you-go pensions, trustees and undergraduates were posed two equivalent questions: one asked respondents to estimate the proportion of over 65 year olds in the year 2025 in percentages, while the other asked respondents to do the same as a ratio. In essence, these questions tested whether respondents were consistent in their responses and whether there was a difference in consistency between the two groups. It was hypothesized that trustees would be more consistent, because the context of their roles and responsibilities is intimately related to the issue.

Figure 3 summarizes the results of this test. The UK government estimates that over 65 year-olds will be 22% of the population in 2025. From Figure 3, it is clear that many trustees were not aware of this estimate and provided much higher estimates. But they did so moderately consistently – that is, the high parametric correlation (Pearson) was 0.71 with a clear linear relationship between percentage and ratio estimates, whatever the trustee estimates of the retired population. Likewise, undergraduates showed no awareness of the government estimate and provided an even wider range of estimates. But they did so inconsistently – that is, the highest correlation (Pearson) was just 0.2 between percentage and ratio estimates. We conjecture that differences in the range of estimates, and the significant difference between the two groups in the level of consistency, are due to the roles and responsibilities of trustees (that is, the relevance of the issue to trustees).

A more complex test of consistency was based upon respondents’ knowledge of the causal relationship between immigration and the funding of pay-as-you-go (PAYG) state pensions. Respondents were asked to estimate the number of young immigrants needed over the next ten years to ‘support adequate state pensions’ based upon their estimates of the proportion of the UK population over 65 years in 2015. Academic treatises on the sustainability of PAYG state pensions demonstrate the

![Figure 3. Consistency of trustees and undergraduates in estimating the proportion of the UK population over 65 years of age in 2025 (where ratios are converted into percentages)](image)

3a. Trustees
Pearson’s correlation coefficient: 0.71
Sign. = 0.024; df = 33; n = 34

3b. Undergraduates
Pearson’s correlation coefficient: 0.2
Sign. = 0.113; df = 63; n = 64

Even after excluding the extreme data points (Figure 3) the correlation was still just 0.3.
relationship between the contributions of the working population and the pension benefits paid to retirees, assuming a constant value of benefits and no additional taxes on the working population (Clark, 2003). It has been also suggested that there are limits on the tax burden that can be imposed on the working population, given global competition for labour and capital investment (Tanzi and Schuknecht, 2000). Increased rates of long-term immigration that contribute to a growing working population could partially fund the rising burden of PAYG state pensions without necessitating increased taxes or decreased benefits. This is an argument of causal logic, although it is the subject of debate amongst academics (Gillion et al., 2000).

Immigration is also a very contentious issue in many western countries (Kreiger, 2005). In the recent UK general election (2005), just after the data collection phase of the NAPF project, the Conservative Party made immigration one of the crucial issues, distinguishing it from the governing Labour Party. We expected a positive relationship between respondents’ projected UK over 65 population and the volume of immigration needed to sustain PAYG pensions. But this correlation was not found for either trustees or undergraduates. In part, this may be because neither group had knowledge of the projected 2025 UK over 65 population. It may also be the case that neither group had a common understanding of the current and 2015 required rates of immigration in the context of widespread disagreement over the proper course of action with respect to funding UK citizens’ retirement income (as indicated by the reception of the Pensions Commission 2005 report; see Clark, 2006).

A number of trustees indicated that no young immigrants were needed to sustain PAYG state pensions; this seems to be a political belief as much as a measured judgement. By contrast, few undergraduates identified 0 per annum or 15,000 per annum as the needed annual level of young immigrants when compared to the trustees. Their estimates were much higher (results available from authors). By framing these problems of conceptual consistency in ways that impinge upon political debate, the context for decision-making may have shifted to include issues otherwise excluded through synthetic or idealized problem sets. Equally, perhaps the judgement of young undergraduates reflected the content of their formal education, whereas trustees’ judgement on this issue reflected their age and political interest in the resolution of the immigration debate.

Implications and conclusions

In this paper, we tackled the question of whether a select group of UK pension fund trustees could be said to be consistent in their decision-making with respect to problems relevant to their roles and responsibilities. As a point of comparison, a group of Oxford undergraduates were asked to solve the same problems providing a means of judging whether trustees were more or less consistent than the undergraduates in decision-making. The empirical strategy was based upon a set of recognized problems and testing protocols drawn from the psychology literature that allows reference back to the results of other researchers.

In the first instance, respondents were asked to solve three inter-related conjunction problems. Responses were compared across each problem, across each group
of respondents, and against the characteristics of the trustees. While a similar proportion of trustees and undergraduates was consistently incorrect in their solutions to these problems, trustees were more consistently correct across the three problems. In the second instance, respondents were asked to solve three problems that required the same decision-technique but posed in different ways. Again, it was found that trustees were more consistently correct than undergraduates, especially in those problems immediately relevant to trustee roles and responsibilities. In the third instance, respondents were asked to solve a set of problems that required knowledge of Bayesian reasoning. Here, it was shown that neither group was consistent decision-makers. They appeared not to recognize the nature of the problem, the best technique for solution, or the fallacies of relying upon one source of information.

At the heart of the analysis of trustee decision-making consistency were two related arguments (or hypotheses): first, following Tversky and Kahneman and others, the ‘norms, habits, and personal characteristics’ of trustees matter when calibrating the degree of trustee consistency; and, second, following Simon and others, the context in which decisions take place matters when assessing the nature and extent of decision-making consistency. At issue was the degree to which trustee education, professional qualifications, and training were related to decision-making consistency. Also at issue was the degree to which trustee roles and responsibilities with respect to the funding of pension benefits could be shown to be related to trustee consistency (compared to undergraduates). It was shown that trustee education and professional qualifications were related to the degree of trustee decision-making consistency. Training was relevant but less important. It was also shown that trustees were more consistent than undergraduates when asked to solve a problem that drew upon specific knowledge derived from the context of their roles and responsibilities.

Notice, however, when asked to solve problems that depended upon specific knowledge of demography and PAYG state pension funding, both groups of respondents recorded a wide range of responses. Many trustees do not have good information about UK demographic projections, although the range of their guesses was less than the range of undergraduate guesses about the same matter. When expressed in a different way, trustees were at least consistent in their guesses, unlike undergraduates. But when asked to match the rate of immigration to the funding of PAYG state pensions, a task that requires a conceptual understanding of the causal relationship between the volume of the working population and the provision of retirees’ pension benefits, it appears that the ‘context’ of decision-making expanded well beyond trustees’ roles and responsibilities to include normative issues such as the proper level of UK immigration. Surprisingly, undergraduates seemed to have a better understanding of this relationship.

These results must be treated with caution considering the small sample of trustees and funds studied. Specifically, we have not been able to match the scope and size of recent industry studies of trustee attitudes, for example. At the same time, there are reasons to be cautious of the significance of attitudinal surveys just as there are reasons to be cautious of the sources of survey participants. Furthermore, if the sample was small, it is not so unusual when compared to recent UK government
studies and current practice in the psychology literature (see Kovalchik et al., 2005). This does not mean, of course, that we should be content with the sample size; more research on the lines suggested is obviously warranted. Nonetheless, our results have significant implications for the governance of pension funds.

- Having shown that education, professional qualifications, and training (task-specific skills) are important for consistent trustee decision-making, attention should be paid to the selection of trustees to serve on pension funds. While the UK government has not mandated formal qualifications for serving trustees, trustee selection according to education and professional qualifications could add to fund decision-making consistency.
- Given the range of trustee task-specific skills even amongst the relatively small number of larger pension funds that participated in the project, attention should be paid to trustee roles and responsibilities. Some trustees may not have the demonstrated capacity or the formal qualifications to be consistent or near-consistent in decision-making that relates to problems of investment strategy and management. Since the choice between competing investment options is made in the context of risk and uncertainty, trustees could be assessed for their fit with different types of roles and responsibilities.
- Given the limits of trustee competence, especially in problems that require sophisticated statistical knowledge and experience, pension funds (and other similar types of institutions such as corporate boards of directors) could consider complementing their decision-making bodies with independent experts capable of contributing to these kinds of judgements. This may be especially appropriate at the level of sub-committees, such as those responsible for fund investment strategies.
- The evidence suggests that the average trustee may not have the education and professional qualifications to play the role of informed consumer of financial advice and decision-making (as suggested in the 2004 Pensions Act). The evidence suggests that training and experience may not make-up the difference if trustees come to investment responsibilities without sufficient qualifications. Again, this suggests that trustees should delegate this type of responsibility more than they do, just as The Pensions Regulator should consider other ways of introducing expertise (not just competence) into pension fund governance.

Of course, trustees play a variety of roles on pension fund boards. Not all need be directly responsible for investment strategy, and not all need have the level of formal qualifications necessary to solve the most complex statistical and conceptual problems. But in that case, there need to be guidelines about roles and responsibilities that recognize differences between trustees in their competence and consistency. Likewise, there needs to be a clear understanding about the roles and responsibilities of trustees as representatives of beneficiaries’ interests as opposed to being active decision-makers responsible for fund investment and strategic tasks and duties. Representation of stakeholder interests on pension fund boards may be best accomplished by a more articulated and developed system of governance and delegation.
The results reported here also bear upon a commitment by governments to shift, in part, responsibility for planning retirement income to individuals. This policy position has been prompted by a number of concurrent economic and political processes extant in modern societies, including the declining relevance of defined benefit pension plans in the private sector and a deeply embedded presumption in favour of individual responsibility, when set in opposition to government or employer paternalism (Strauss, 2006). Whatever the sources promoting greater individual responsibility, many governments acknowledge that realizing viable retirement incomes in the future depends on citizens’ competence and consistency. It is not surprising that governments have promoted financial literacy and the use of idealized decision-trees to encourage better-quality decision-making. The results reported here and elsewhere, however, for neophytes and those responsible for the pension welfare of others are hardly a ringing endorsement of average competence and consistency.

That formal education and professional qualifications have a significant role to play in judgement should be reason to qualify public confidence in pension regimes that rely entirely or in part upon individual competence and consistency. No doubt, some citizens will thrive with greater responsibility, but others will not, particularly if asked to adapt and respond in ways that test their underlying knowledge and understanding of economic and demographic prospects. We have less confidence than some in the competence and consistency of average citizens.

References


Doherty, M. E. and Mynatt, C. (1990) Inattention to P(H) and to P(D/-H): a converging operation. Acta Psychologica, 75:


