Compressed environments: Unbounded optimizers should sometimes ignore information

Berg, Nathan and Hoffrage, Ulrich

2010

Springer

Compressed Environments: Unbounded Optimizers Should Sometimes Ignore Information

Nathan Berg · Ulrich Hoffrage

Abstract   Given free information and unlimited processing power, should decision algorithms use as much information as possible? A formal model of the decision-making environment is developed to address this question and provide conditions under which informationally frugal algorithms, without any information or processing costs whatsoever, are optimal. One cause of compression that allows optimal algorithms to rationally ignore information is inverse movement of payoffs and probabilities (e.g., high payoffs occur with low probably and low payoffs occur with high probability). If inversely related payoffs and probabilities cancel out, then predictors that correlate with payoffs and consequently condition the probabilities associated with different payoffs will drop out of the expected-payoff objective function, severing the link between information and optimal action rules. Stochastic payoff processes in which rational ignoring occurs are referred to as compressed environments, because optimal action depends on a reduced-dimension subset of the environmental parameters. This paper considers benefits and limitations of economic models versus other methods for studying links between environmental structure and the real-world success of simple decision procedures. Different methods converge on the normative proposition of ecological rationality, as opposed to axiomatic rationality based on informational efficiency and internal consistency axioms, as a superior framework for comparing the effectiveness of decision strategies and prescribing decision algorithms in application.

N. Berg (✉)
School of Economic, Political, and Policy Sciences (EPPS), University of Texas at Dallas,
GR 31 211300, Box 830688, Richardson, TX 75083-0688, USA
e-mail: nberg@utdallas.edu; prof.berg@gmail.com

U. Hoffrage
Faculty of Business and Economics, University of Lausanne, Batiment Internef,
CH-1015 Lausanne, Switzerland
e-mail: ulrich.hoffrage@unil.ch

Introduction

This paper offers a formal decision-theoretic result showing a less-is-more effect regarding the information requirements of expected-payoff-maximizing decision algorithms. In ongoing debates over normative criteria for what it means to make good, or rational, decisions, economists and formal decision theorists tend to put forward axiomatic criteria, such as transitive preferences or the Savage Axioms that underlie expected-utility theory, as appropriate normative benchmarks (Raiffa 1986). Axiomatic notions of rationality require self consistency above all else, rather than absolute levels of performance, and their requirements of self consistency apply, by definition, across all decision domains. In contrast, ecological rationality focuses on the match between a decision algorithm and the environment in which it is applied: if the match is good enough that the decision maker avoids harm, then the algorithm is likely to persist. Because of the high dimensionality of the decision maker's action space, constant change in the stochastic payoff environment, and profound limitations on any mind or machine's ability to list future outcomes let alone assign probabilities to those outcomes, proponents of ecological rationality argue that there is little adaptive pressure in most natural environments to optimize (Gigerenzer et al. 1999).

Optimization may be well-defined but infeasible, as in the game of chess. Or optimization may be ill-defined, as in mate choice—who, after all, could list the criteria for optimal selection of his or her spouse and evaluate potential spouses by these criteria? Whether infeasible or ill-defined, axiomatic rationality is inapplicable, and ecological criteria requiring only that decision processes work well enough to survive would seem to provide more practical normative content.

One major disagreement in debates about axiomatic versus ecological rationality, with clear ramifications in applied problems of designing machines to monitor signals and return judgments, classifications or decisions, concerns the quantity of information. Is more always better? Game theorists acknowledge that, in some strategic settings, possessing information can be disadvantageous (say, when one is a witness to a crime and consequently becomes targeted because of possession of unwanted information). They claim, however, that more information is necessarily better in games against nature (Raiffa and Shlaifer 1961). There is also a sizeable literature in psychology showing that human subjects often ignore relevant information, with the interpretation that ignoring is pathological (Kahneman et al. 1982).

This paper provides a formal model in which it is possible to state precise conditions under which expected-payoff-maximization is consistent with ignoring relevant information. If expected-payoff maximization is consistent with ignoring relevant, or predictive, information, then there can be no competitive pressure selecting for decision rules that take more information into account.

As background, consider the finding that sleep deprivation, even at moderate levels, increases the risk of an auto accident by an amount similar to that of alcohol

intoxication (Mahowald 2000). Time of day also correlates strongly with occurrences of auto accidents (Clarke et al. 2002; Dobie 2002; Pack et al. 1995). And geographic location is another factor known to condition chances of a traffic accident (Mohan 2001). Does it follow, then, that rational decisions about whether to use seat belts should necessarily depend on these variables because they predict outcomes that affect drivers' payoffs?

This question is also relevant for machine learning and for the development of algorithms used by machines: Would the expected-payoff-maximizing decision algorithm used by a machine programmed to decide on seat belting necessarily depend on location, time of day, and the previous night's sleep? Or might optimal decision rules rationally ignore some of that information? This paper shows that the answer to the last question is affirmative. Expected-payoff-maximizing decision algorithms rationally ignore signals that are correlated with payoffs if the stochastic payoff environment satisfies several properties.

According to the cost-benefit model implicit in the objective of maximizing expected payoffs, optimal seat belt decisions should use information about risk factors for traffic accidents that vary through time and space, as well as the condition of the driver, together with a driver's risk attitudes, subjective costs of seat belting, expected fines for driving without a seat belt, and the intensity of enforcement policies regarding seat belt rules. Of course, any particular decision, either for or against seat belting, can be rationalized within the expected-payoff maximization framework. Given sufficiently extreme dislike of seat belts or low enough risk parameters, the decision to forgo the proven safety benefits of seat belts is consistent with maximization. More generally, drivers driving at variable times, locations, and with varying amounts of rest will maximize expected payoffs by opting for seat belts in a contingent manner, based on a seat-belt-decision algorithm that scores environments according to a weighted index of risk factors. Many drivers in the real world do not do this, however.

Many drivers in the real world instead wear their seat belts automatically, regardless of time, place or their own physiological condition. Is such behavior irrational? No. Habitual or automatic decision processes that disregard observable signals, which help predict the chances of an accident, are perfectly consistent with expected-payoff maximization, so long as drivers regard the net benefits of seat belting as positive in all environments, including those with the lowest possible risk. In other words, if the decision maker has cost-benefit and risk parameters for which it is optimal to wear the seat belt in every driving environment, then the expected-payoff-maximizing decision algorithm will ignore signals that are correlated with payoffs, contrary to the standard normative message in most of economics and decision theory suggesting that more information is better, and that ignoring relevant information is pathological.

Rather than an isolated case, optimizing seat belt strategies that ignore statistically valid predictors of accidents are one manifestation of a general principle. Bullock and Todd (1999) identified forms of asymmetry in the external environment favoring information-frugal decision making. Forster (1999) suggested systematic relationships between the structure of the environment and success of simple decision rules. The robustness of simple algorithms with low informational

requirements was demonstrated in a variety of contexts (Simon 1982; Bookstaber and Langsam 1985; Brooks 1991; Martignon and Schmitt 1999; Todd 1999; Martignon et al. 2008; Gigerenzer and Brighton 2009). Goldstein and Gigerenzer (2002) analyzed cue-criterion correlations that underlie a special class of less-is-more effects. Hogarth (2001) and Hogarth and Karelaia (2005) specified conditions on data generating processes that favor frugal rules of inference.

In spite of these developments, our understanding of the links between environmental structure and the effectiveness of simple decision procedures remains incomplete. This paper proposes an economic model to generalize the question: What conditions on environments enable low-information decision rules to match the performance of decisions based on full optimization? We approach this question in steps. What is an environment? What does low-information mean? And with respect to what metric is performance to be measured? After formalizing the requisite concepts and introducing a rigorous definition of the decision-making environment, the paper illustrates advantages of the formal modeling approach using a particular example with special functional forms. An easy-to-interpret analytic condition results, specifying precisely when payoffs and probabilities interact to produce optimal action rules that are independent of objectively predictive decision cues.

The General Decision Problem

Let Y be a continuous or discrete random variable representing states of the world that are ex ante unobservable to the decision maker. The vector of cues X represents observable environmental factors used to form expectations about Y. Assumption 1 rules out redundancy, or perfect collinearity, among cues:

A1   (non-redundancy) $E[XX']$ exists and is full rank.

The trivial case of cues that fail to predict states is also ruled out by requiring each component of X to be state-relevant. To be state-relevant, there must be at least one pair of values in the cue's support at which the conditional expectation of Y is non-constant with respect to X:

A2   (state-relevancy) $E[Y|X]$ is non-constant in each component of X for some value on its support.

The decision variable labeled a (for action) takes on values in the action space $\mathcal{A}$. The function $f_{Y|X}(Y, X, a)$ denotes the conditional density of Y given X, which may or may not depend on a. The payoff function $\pi(Y, X, a)$ serves to rank conditional distributions of Y according to the expected payoff criterion.

Mappings into Outcomes and Payoffs

Figure 1 illustrates three different mappings from observable information (X) and actions (a) into outcomes (Y) and payoffs (π). Case 1 is the simplest, corresponding to tasks frequently studied in experiments with human subjects such as paired
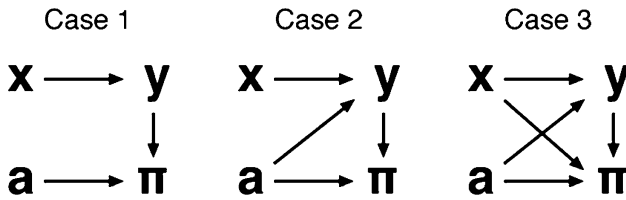
Fig. 1 Mappings from cues (X) and actions (a) into states(Y) and payoffs (π)

comparison. In one well-known paired comparison task, two German cities are drawn at random from a fixed list of cities, and subjects are asked to state which among the two has a larger population (Gigerenzer et al. 1991). The state Y, unknown to the mind or machine faced with the comparison task, codes whether the larger city is listed first or second: $Y = 1$ or 2. Unconditional probabilities o the events $Y = 1$ and $Y = 2$ are fixed at 1/2 by the experimental design. Obviously, whether the decision maker picks the first or second city ($a = 1$ or $a = 2$) has no effect on the frequency distribution of Y, indicated by the absence of an arrow connecting a to Y in Fig. 1, Case 1. Choices of a together with Y do, nevertheless, jointly determine payoffs π. The combinations $(a, Y) = (1, 1)$ and $(2, 2)$ correspond to correctly picking the first or the second city, respectively, while $(1, 2)$ and $(2, 1)$ represent incorrect paired comparisons.

What makes the comparison task interesting is the possibility of using information in X by exploiting its components' correlations with Y. In the city example, X records city characteristics, correlated with Y to varying degrees, such as presence of professional soccer teams, train stations and universities. Correlation between X and Y reflected by arrows connecting those two symbols is a constant feature of the model in all three cases in Fig. 1. Absence of an arrow between X and π in Case 1 implies that payoffs depend only on actions and outcomes, not on other observable covariates. Most paired comparison tasks studied in the psychology literature satisfy this condition, in that correct answers are rewarded uniformly, regardless of other contextual information contained in X. For example, payoffs are the same for correctly answering which of two cities has the larger population, regardless of whether both cities have train stations or not.

Cases 2 and 3 in Fig. 1 introduce additional complexity that requires further examples to illustrate. Suppose now that states (Y), or outcomes, represent low and high crop yields; a is choice of fertilizer types; π maps (Y, a) combinations into payoffs as measured by the cash value of crop yields net of costs; and X codes observable weather conditions at planting time. The arrow in Case 2 from a to Y reflects the fact that choice of fertilizer affects the probabilities of high and low crop yields. Unlike paired comparison tasks in which the actor's decisions do not influence experimentally determined probabilities of outcomes, the entire motivation for acting in many real-world environments is to improve the chances of a good outcome.

In Case 3 of Fig. 1, an arrow links X and π, representing the most complicated case. Extending the crop yield example of the previous paragraph, Case 3 would be applicable if weather conditions coded in X also influenced thecash value of any

realized level of crop yield Y. This would be the case if pre-planting weather conditions simultaneously predicted low yields and higher than average market prices at harvest time. In such a case, X conditions Y, as usual, but also moves π up and down for any fixed value of Y.

Definition of Environment

Utilizing the framework introduced above, one can make a precise definition of the decision-making environment. Suppose there exist states Y, cues X, and actions a. We say that the pair of functions $\epsilon \equiv \{f_{Y|X}(Y, X, a), \pi(Y, X, a)\}$ is an *environment*, where $f_{Y|X}$ is the conditional density of Y given X, and π is a scalar valued payoff function representing fitness rankings, by the expected utility principle, over conditional distributions of Y.

Ecological Versus Axiomatic Rationality

Formal specification of the environment usefully delimits the generalizability of comparisons between two action functions, algorithms or decision procedures. Consider two action functions $a_1(X)$ and $a_2(X)$, and two environments ε_A and ε_B. It often happens that what works well in one setting is ineffective in another. Such context-dependent evaluation of decision procedures can be translated as: $a_1(X) \succ_A a_2(X)$ and $a_2(X) \succ_B a_1(X)$. In other words, $a_1(X)$ performs better than $a_2(X)$ in environment A, but the reverse is true in environment B.

From the perspective of ecological rationality, the performance of algorithms is ε_i-specific (where i indexes a set of environments), based on rankings of expected payoffs in environment ε_i. Action $a(X)$ is adaptive, or ecologically rational with respect to ε_i if it maximizes expected payoffs in ε_i for some set of action rules under consideration.

A famous example is Simon's (1982) notion of intelligent search. In environments with food uniformly distributed on a plane, intelligent search, supported by capacities such as vision and memory, is equivalent to random search based on nothing more than random movement in the plane. In environments where food is distributed in clusters, intelligence, or rational foraging, means something quite different, and memory can be exploited for adaptive gain. From the perspective of ecological rationality, it is a mistake to regard memory as a beneficial cognitive capacity in a universal sense. Rather, memory helps in some particular environments, but there are also environments where forgetting is advantageous, for instance, when nonrandom forgetting improves the accuracy of inferences (Schooler and Hertwig 2005), or when it enables forgiveness or peace of mind.

In much of the judgment and decision making literature, predictive accuracy is tacitly assumed to measure performance, which would mean that π simply counts the number of correct decisions, or is an increasing function of that count. Connolly (1999) argues that predictive accuracy has special philosophical status in identifying causal links. But important counter examples with highly asymmetric losses may warrant reconsideration of predictive accuracy as a proper normative benchmark. In Bullock and Todd's (1999) mushroom selection task, a decision rule that maximizes

predictive accuracy would aim to not eat mushrooms whenever they are poisonous, and to eat mushrooms whenever they are non-poisonous. Compared to this predictive-accuracy-maximizing strategy, more conservative strategies that waste more non-poisonous mushrooms, but do better at avoiding poisonous mushrooms, are easily superior in terms of promoting survival of the species. In the case of seat belting, the predictive-accuracy-maximizing strategy would maximize the overall chance of either wearing a seat belt when accidents occur, or of not wearing a seat belt when no accident occurs. Again, because of asymmetric losses, departing from predictive-accuracy maximization may be adaptive. Because smart behavior does not necessarily maximize predictive accuracy, it is therefore important to be explicit about π, any asymmetries of costs and benefits, and defend predictive accuracy on a case-by-case basis.

In contrast, the axiomatic approach (e.g., Luce and Raiffa 1957) promotes the position that there are appealing normative guidelines for ranking $a_1(X)$ and $a_2(X)$ which hold universally across all environments, obviating the need to specify ε. Demonstrations of context specificity, such as Simon (1982), in which transitivity and virtually all other imaginable constraints on decisions in the uniformly distributed food environment fail to characterize effective search rules, would seem to make the axiomatic approach less attractive. Following Simon's lead, Todd (2000), Gigerenzer and Selten (2001), Berg (2005) and Berg and Hoffrage (2008) demonstrate high-stakes instances in which the meaning of *adaptive* is clearly context specific. Negating the axiomatic approach to defining rationality does not imply relativism, however. Indeed, examples of non-axiomatic theories of decision making abound. The challenge is to identify observable features of environments that predict when simplicity will succeed.

Expected-payoff Maximization

Once the environment is specified, making π explicit, the question of determining how well different decision rules perform is far from obvious. Because there is random chance in the environment, every choice of an action gives rise to a probability distribution of payoffs. Choosing the action rule with the highest associated expected payoff is perhaps the most common benchmark, but many other statistics of the payoff distribution other than the mean may be important. Linear regression and Bayesian prediction algorithms are commonly used benchmarks in simulation studies that seek to gauge the performance of simple heuristics using real-world data (Czerlinski et al. 1999; Martignon and Schmitt 1999; Martignon and Hoffrage 2002). Those benchmarks provide useful tools for measuring the predictive value of information, for example, the predictive gains of moving from a decision rule based on a single cue to one that utilizes ten pieces of information.

In the abstract environment of the economic model, the unconstrained expected-payoff maximizing decision rule, assuming it exists and is unique, searches over every possible rule for mapping K pieces of observable information, X, into actions. The expected payoff maximizer of $E[\pi(Y, X, a)|X]$ with respect to a is, in general, a function of all K cues and is denoted $a^*(X)$.

Defining Frugality

Suppose the decision maker faces no cognitive constraints, no information costs, and has a complete understanding of the data generating process (i.e., the joint frequency distribution of Y and X) that determines future outcomes. Under what conditions on $\{f_{Y|X}(Y, X, a), \pi(Y, X, a)\}$ is the link between the elements of X and $a^*(X)$ broken, despite the dependence of E[YlX] on all elements of X?

The decision rule $a(X)$ exhibits one basic form of frugality whenever it depends only on a strict subset of the K elements in X. In other words, there are K predictors of payoff-relevant outcomes, but the expected-payoff-maximizing algorithm depends on a strict subset of them. There are other simple rules in the literature, such as Gigerenzer and Goldstein's (1996) lexicographic take-the-best strategy, that depend on a strict subset of the cues most of the time, but occasionally look up all K cues. To deal with such cases, frugality can be defined as a statistic of the stochastic number of cues used in the decision procedure. For example, sometimes the algorithm requires all 10 pieces of information, sometimes only 1, but on average, it looks up 3 cues. An obvious frugality measure, then, is 1 minus the average fraction of K cues used. By this measure, frugality ranges over the unit interval, with maximally frugal decision rules, which never look up any cues at all, indicated by a frugality measure of 1. A decision rule with positive frugality is referred to as frugal. Of course, there are many frugal decision rules, most of which are maladaptive. The important question is whether one can identify specially structured environments in which the maximally adaptive rule $a^*(X)$ is guaranteed to be frugal according to the definition above.

A caveat is in order. Frugal decision rules are not necessarily heuristics, in the sense of Gigerenzer and Todd (1999). Heuristics are frugal, but the converse does not hold. The economic, or top-down, approach advanced in this paper is well suited for identifying conditions under which adaptive decision procedures have low information requirements. It stops short, however, of discovering new heuristics. To move from the set of frugal, adaptive decision rules to its strict subset comprised of heuristics would require additional specificity with regard to the mechanisms of the human mind. Additional psychological theory and empirical evidence are required. Still, valid generalizations derived from economic models linking environmental structure to frugal and adaptive decision rules must apply to heuristics as well, and therein lies the opportunity for the economic approach to provide insight. Understanding when reduction of dimensionality is possible—that is, learning to recognize *compressed environments*—is itself of tremendous value, as statisticians (Hansen and Bin 1996) and, increasingly, economists (Zellner et al. 2001) have shown.

Compressed Environments

The environment ε_θ is said to be *compressed* whenever its expected-payoff-maximizing decision algorithm is frugal. In such cases, a strict subset of the relevant information permits optimal performance and, in this sense, its stochastic payoff structure is compressed.

Next, we present a model of a one-cue environment in which the condition guaranteeing compression (i.e., that $a^*(X)$ ignores X) can be stated explicitly. Compression is equivalent to rational ignoring. Its source is the interaction of payoffs and probabilities in the external environment that allow optimal action to get by with reduced informational requirements. The advantage of the economic approach is that the precise interaction of probabilities and payoffs that leads to compression can be seen directly in algebraic form, and its sensitivity to various parameters of interest may be easily examined.

An Example: Theft Deterrence in a Small-scale Society

Consider a small-scale society with boat-making technology and a high degree of dependence on boats and boat operators for survival. The decision problem is to choose the quantity of resources to be allocated to theft deterrence, which includes any activity that reduces the chances of the boat being stolen. Time and material resources allocated to theft deterrence are aggregated into the scalar action variable a, ranging from zero to the maximum effort 1, (i.e., $0 < a < 1$). Theft deterrence includes such activities as nighttime patrolling, camouflaging, and physically securing the boat with anchors and ropes, which incurs non-negligible and increasing marginal cost with each increase in effort, a.

States of nature encode possession or loss of the boat. Theft is coded as the event $Y = 0$ (i.e., absence of the boat), and no-theft is coded as $Y = 1$ (i.e., possession of the boat). There is just one observable cue X, which correlates positively with Y and therefore helps predict boat theft. The cue takes on values of 0 and 1 depending on whether the group was attacked in the preceding week, coding the missing or present state of group members lost in the attack:

$$X = \begin{cases} 0 & \text{if group members are missing, killed in attacks by rival groups,} \\ 1 & \text{if all group members present.} \end{cases} \tag{1}$$

Loss of group members tends to coincide with boat theft because the two kinds of losses share a common cause, which is the presence of hostile competitors in the vicinity. Thus, the worst and best joint outcomes, theft and attack (0,0) and no-theft and no-attack (1,1), are more likely than joint outcomes with intermediate payoffs, theft and no-attack (0,1) and no-theft and attack (1,0).

The motivation for groups to allocate effort to theft deterrence is that higher choices of a reduce the probability of theft, captured by conditional probabilities of theft that are decreasing in deterrence effort with the following functional forms:

$$\Pr(Y = 0 \text{ (theft)} \mid X = 0 \text{ (attack)}) = f_{00}(1 - a), \tag{2}$$

$$\Pr(Y = 0 \text{ (theft)} \mid X = 1 \text{ (no attack)}) = f_{01}(1 - a). \tag{3}$$

Given that group members were killed and zero deterrence is applied ($a = 0$), Eq. 2 indicates that the conditional chance of theft is f_{00}. As more resources are spent on deterrence, the chance of theft conditional on group members having been killed is reduced, shrinking to zero in case maximum effort is applied ($a = 1$). Similarly,

Eq. 3 states that, conditional on no recent loss of group members due to hostilities, the chance of theft is f_{01} for zero deterrence, decreasing to zero in case maximum deterrence is applied. The assumption $f_{00} > f_{01}$ captures the idea that, for any choice of a, the probability of theft is greater if group members have been attacked than if no attack occurred.

The opportunity cost of resources allocated to a is assumed to be an increasing, convex quadratic function, reflecting the idea that doubling the level of deterrence more than doubles its cost. The payoff function is therefore specified as:

$$\pi(Y, X, a) = \pi_{YX} - ca^2/2, Y, X = 0, 1. \tag{4}$$

The four constants π_{YX} shift payoffs up and down across the four joint state-cue combinations (Y, X), and the cost parameter c scales the quadratic costliness of action a.

We assume the group is always better off in possession of the boat ($Y = 1$ vs. $Y = 0$) whether group members have been killed or not ($X = 0$ or 1):

$$\pi_{11} > \pi_{01} \text{ and } \pi_{10} > \pi_{00}. \tag{5}$$

Similarly, we assume that the group is always better off with more of its members alive ($X = 1$ vs. $X = 0$) whether its boat is stolen or not ($Y = 0$ or 1):

$$\pi_{11} > \pi_{10} \text{ and } \pi_{01} > \pi_{00}. \tag{6}$$

By assumption, the boat can be operated more efficiently with a full set of boat operators. Therefore, because the boat operates less efficiently with fewer operators, the value of the boat is greater when all group members are alive than when some are killed:

$$\pi_{11} - \pi_{01} > \pi_{10} - \pi_{00}. \tag{7}$$

The universe of theft-deterrence environments is parameterized by seven values, stacked into vector form as $\theta = [\pi_{11}, \pi_{01}, \pi_{10}, \pi_{00}, f_{01}, f_{00}, c]$. The two conditional probabilities must lie within the unit interval and respect the inequality $f_{00} > f_{01}$. Payoffs must satisfy inequalities Eq. 5 through Eq. 7. The cost scaling factor c must satisfy $c \geq \max\{f_{00}(\pi_{10} - \pi_{00}), (1 - f_{01})(\pi_{11} - \pi_{01})\}$ to guarantee existence of an interior maximizer with respect to a. The boundaries of these restrictions on θ define a dense subset of 7-dimensional Euclidean space in which each point corresponds to an admissible value of θ and, thus, to one particular realization of the theft-deterrence environment.

The expected payoff function is:

$$\mathrm{E}[\pi(Y, X, a)|X] = \begin{cases} (1 - f_{00})\pi_{10} + f_{00}\pi_{00} + af_{00}(\pi_{10} - \pi_{00}) - ca^2/2 & \text{if } X = 0, \\ (1 - f_{01})\pi_{11} + f_{01}\pi_{01} + af_{01}(\pi_{11} - \pi_{01}) - ca^2/2 & \text{if } X = 1. \end{cases} \tag{8}$$

Without further restrictions, it is clear from the piecewise definition of the expected payoff function that, depending on whether $X = 0$ or 1, different actions will be required to maximize expected payoffs. Indeed, the expected payoff maximizing action rule is itself a piecewise-defined function that provides different levels of optimal effort depending on the observed value of X:

$$a^*(X) = \begin{cases} f_{00}(\pi_{10} - \pi_{00})/c & \text{if } X = 0, \\ f_{01}(\pi_{11} - \pi_{01})/c & \text{if } X = 1. \end{cases} \tag{9}$$

There is a condition on θ, however, under which cancellation of payoffs and probabilities occurs, and $a^*(X)$ is consequently independent of X. Thus, the theft-deterrence environment θ is compressed if:

$$\frac{\pi_{10} - \pi_{00}}{\pi_{11} - \pi_{01}} = \frac{f_{01}}{f_{00}}. \tag{10}$$

Any theft-deterrence environment satisfying Eq. 10 gives rise to a payoff-maximizing effort rule that optimally ignores X. Action functions in compressed environments choose levels of theft deterrence without regard to observed hostile attacks, even though hostile attacks correlate with and are indeed causally linked to boat theft. The intuitive explanation for this is that, conditional on loss of group members, the decrease in the value of the boat given by the ratio $\frac{\pi_{10} - \pi_{00}}{\pi_{11} - \pi_{01}}$ is exactly offset by an increase in the conditional risk of theft given by $\frac{f_{00}}{f_{01}}$. In compressed environments, the two ratios measuring decrease in the boat's value on one hand, and increase in the risk of theft on the other, are reciprocals. Thus, compressed environments are characterized by inverse movement of payoffs and probabilities across cue observations, $X = 1$ versus $X = 0$. Inversely moving payoffs and probabilities produce a flat expected payoff function even though frequencies and payoffs are, themselves, nonconstant with respect to X.

Visualizing Cancellation in the Space of Environments

A single realization of the theft-deterrence model requires seven parameter values for the elements in the vector θ, too many to visualize easily. This section introduces a simplified re-parameterization of payoffs and probabilities in terms of ratios that reflect values conditional on attack $X = 0$ versus no-attack $X = 1$. The simplified re-parameterization facilitates an intuitive visualization of the model, within which the set of compressed environments can easily be seen.

First, the re-parameterization. The value of the boat given no-attack, $X = 1$, is $\pi_{11} - \pi_{01}$. Similarly, the value of the boat given an attack is $\pi_{10} - \pi_{00}$. We define the ratio of the boat's value conditional on attack to its value conditional on no attack as

$$g \equiv \frac{\pi_{10} - \pi_{00}}{\pi_{11} - \pi_{01}}. \tag{11}$$

If the boat requires a larger number of people to operate than is available after the attack, then the boat loses value following the attack, and g will be <1, with a minimum of zero in the event that the boat loses all value. If, on the other hand, the boat gains value because of the attack, for example, if the boat provides an essential means of escaping further attack, then g is larger than one, with no upper bound. The inequality on payoffs introduced earlier (Eq. 7) translates into g ranging between 0 and 1.

In the parameterization of the previous section, the joint distribution of theft and attack was determined, already in abbreviated form, by two conditional probability parameters, $\Pr(\text{theft} \mid \text{attack}) = f_{00}$ and $\Pr(\text{theft} \mid \text{no attack}) = f_{01}$. Here, we

introduce a single ratio-of-theft-likelihoods parameter l that measures the gross percentage increase in the likelihood of theft, conditional on attack:

$$l \equiv f_{00}/f_{01}. \tag{12}$$

As long as theft and attack are positively correlated, the restriction $f_{00} > f_{01}$ holds, which translates as $l > 1$.

Given the parameters g and l (g capturing changes in payoffs across states of X, and l capturing changes in theft probabilities), we can express a similar ratio of expected-payoff-maximizing actions, that is, optimal effort conditional on $X = 0$ over optimal effort conditional on $X = 1$:

$$a^*(0)/a^*(1) = gl. \tag{13}$$

Any set of environments along which the product gl is constant is referred to as a *iso-action-ratio curve*. Iso-action-ratio curves are subsets of the universe of environments along which the ratio of optimal effort following attack versus no attack is the same. For example, the iso-action-ratio curve given by $gl = 1.2$ indicates all environments where the optimal reaction to a change in signal X, from 1 to 0, is to increase effort by 20%. There is one special iso-action-ratio curve defined by the restriction $gl = 1$ that exactly defines the set of compressed environments. Everywhere this restriction holds, optimal action is the same regardless of which attack signal is observed.

Figure 2 depicts the valid parameter space for the theft-deterrence model projected into (g, l) space. The inequalities on payoffs in previous sections give rise to two vertical lines at $g = 0$ and $g = 1$ that bind g between 0 and 1. The restriction that $g > 0$ follows from the assumption that the boat has positive value. The restriction that $g < 1$ reflects the assumption that the boat's value decreases conditional on attack. The horizontal line $l = 1$ depicts the lower bound on l (i.e., $l > 1$), which follows from the assumption that X and Y are positively correlated.

Both dimensions of Fig. 2 are ratios of values in which the numerator corresponds to $X = 0$ and the denominator corresponds to $X = 1$, with the ratio of boat values given by g on the x-axis, and likelihoods of theft given by l on the y-axis. Figure 2 makes it straightforward to visualize the set of compressed environments, given by the iso-action-ratio curve $gl = 1$, along which rational ignoring occurs. Rational ignoring along the compressed set occurs because percentage increases in the likelihood of theft, l, are the exact reciprocal of percentage decreases in the value of the boat, g.

If the modeler wishes to introduce perceptual limits in action space on how finely levels of effort can be adjusted, a just-noticeable-differences argument augments the set of ignoring environments from the one-dimensional iso-action-ratio curve $gl = 1$ to the dense set in (g, l) space defined as follows. Suppose that, whenever the optimal action rule prescribes action across attack/no-attack ($X = 0/X = 1$) states that are less than $\tau \times 100\%$ different, then the decision maker ignores the signal X and applies the same action in both states. That is, a τ-sensitive decision maker ignores X and chooses the same level of effort regardless of attack whenever

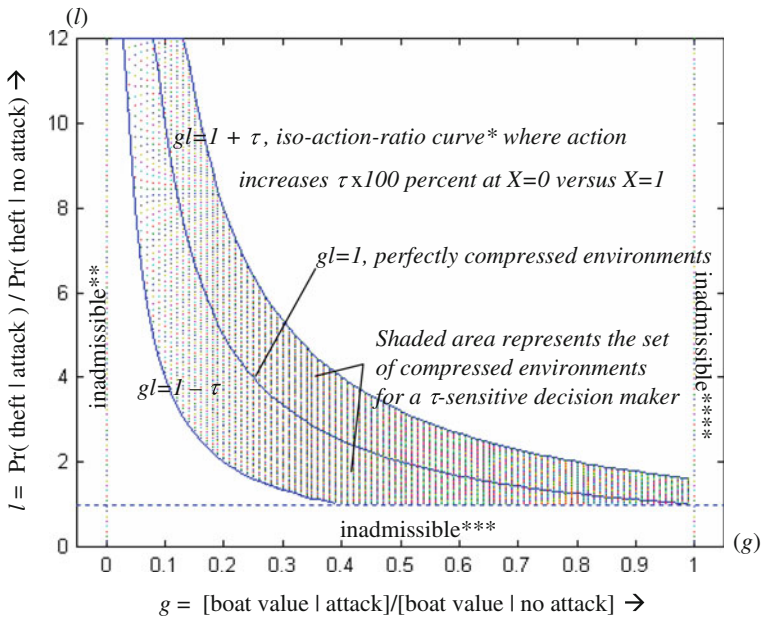$$1 - \tau < a^*(0)/a^*(1) < 1 + \tau. \tag{14}$$

Fig. 2 Compressed environments within the space of al boat-theft deterrence environments. *The action ratio is defined as $a^*(0)/a^*(1)$, the expected-payoff-maximizing action (i.e., optimal level of effort) conditional on the signal $X = 0$ divided by the expected-payoff-maximizing action conditional on $X = 1$. An *iso-action-ratio curve* is a set within parameter space along which the action ratio is constant. For example, the set of environments for which $a^*(0)/a^*(1) = 1.2$ is an iso-action-ratio curve indicating all parameterizations for which a change in the signal X, from 1 to 0, leads to a 20 percent increase in action. **Points to the left of the vertical line $g = 0$ are inadmissible because of the assumption that the boat has positive value, $g > 0$. ***Points below the horizontal $l = 1$ are ruled out by the assumption that theft is positively correlated with attack. ****The restriction that $g < 1$ corresponds to the assumption that the boat's value is reduced in the event of attack. The model easily allows for extension of parameter space by admitting points $g > 1$. However, based on the previous assumption that $l > 1$, there are no compressed environments in the extended region to the right

Given perceptual or technological limitations that make an equivalence class of any action levels that differ in percentage terms by less than τ, the ignoring subset in the space of environments indexed by (g, l) is enlarged, shown by the shaded area in Fig. 2.[1]

There is another large set of compressed environments in which rational ignoring occurs for a different reason, connected to the cost parameter c, than in the set of compressed environments depicted in Fig. 1. Recall that the optimal action rule is given by Eq. 9. An earlier section gave a lower bound on c to insure that $a^*(X)$ takes on admissible values between 0 and 1. If, however:

$$c < \min\{f_{00}(\pi_{10} - \pi_{00}), f_{01}(\pi_{11} - \pi_{01})\}, \tag{15}$$

---

[1] An important theme in the just-noticeable-differences literature is asymmetry of perceptual sensitivity with respect to deviations from the status quo. Allowing for asymmetric τs would add an additional degree of psychological plausibility without changing the main conclusion, namely, that the set of environments in which ignoring is rational is nonempty.

then the cost parameter is so small that the optimal action is maximum effort, regardless of the signal X. In this case X is correlated with payoffs, yet the optimal action algorithm ignores it, but not because of inversely moving payoffs and probabilities (i.e., g and l are not necessarily reciprocals). Rather, compression occurs in this case because action is so cheap that the maximum effort is applied regardless of varying signals about risk. This is analogous to wearing a seat belt every time one drives, ignoring time-varying signals about risks of an auto accident.

Conclusion

This paper makes no categorical claim about benefits of ignoring information. Ignoring is, without doubt, detrimental to the wellbeing of decision makers and the societies they belong to in some cases. However, the model in this paper does show clearly that ignoring is not always detrimental. In the theft-deterrence model, expected payoff maximizing decision rules ignore information that objectively helps predict future states and payoffs when payoff and probabilistic information is compressed in the proper way. The compression condition does not require anything pathological or degenerate about the stochastic payoff process. Rather, it depends on cancellation of probabilities and payoffs that de-links variables, which predict both probabilities and payoffs, from expected-payoff-maximizing action.

Two frequently cited experiments in which predictive information is ignored are Kahneman and Tversky's (1973) engineer-lawyer problem and Tversky and Kahneman's (1982) taxi cab problem. In both studies, respondents were asked to produce posterior probabilities given experimenter-controlled prior probabilities, hit rates, and false-positive rates. The studies showed a surprising lack of sensitivity to changes in base rates, which came to be known as base-rate neglect, or the base-rate fallacy. To the extent that experimental subjects receive higher intrinsic payoffs from correct inference when the probability of success is low (i.e., take greater psychological satisfaction in solving more difficult problems), base-rate neglect could be consistent with expected-payoff maximization. It may be worthwhile therefore to vary monetary payoffs by difficulty, measure the economic losses associated with base-rate neglect, if any, and collect post-experiment survey data to investigate the possibility of asymmetric subjective payoffs in experimental conditions that otherwise appear symmetric.

Another interesting case close to those working in the scientific community concerns payoffs from scientific discoveries and their probabilities of occurrence. Scientific projects with low probabilities of success tend to have large payoffs, whereas projects that are nearly certain to verify scientists' hypotheses are unsurprising and therefore have nearly zero payoff. Perhaps because of cancellation of probabilities and payoffs, many scientists seem to ignore payoffs and probabilities, adopting simple heuristics for choosing research agendas, such as *Follow the line of research my advisor works in* or *Pick projects that are*

intrinsically interesting to me, rather than estimating probabilities of success and choosing projects according to a maximization algorithm.

The normative implications of ignoring are, in general, ambiguous. This stands in contrast to decision-theoretic prescriptions advising, in virtually all contexts, that Bayes rule be followed by conditioning action on the full set of observable predictors. One of the main implications of the demonstrated existence of compressed environments in which rational ignoring occurs is that ecological (i.e., environment-specific) rationality, rather than universal rationality axioms focused on informational efficiency and internal consistency, are needed to properly evaluate individual and aggregate consequences of ignoring information. Less information does not necessarily imply reduced performance.

Using economic models to identify conditions under which informationally frugal action performs just as well as information-greedy optimization is analogous to simulation studies which find that simple decision rules can match (or exceed in out-of-sample prediction) the performance of prediction rules based on regression or Bayesian networks. Bullock and Todd (1999) investigated a similar set of questions using agent-based modeling and computer simulation. As one might anticipate, distinct methods lead to distinct and sometimes difficult to compare results. The simulation approach requires specific distributional assumptions and functional forms, or else relies on particular databases of real-world data sets that are, at some stage of analysis, interpreted as statistically representative of a larger population of environments.

The economic approach, on the other hand, searches a much larger space of possible environments, although still within the confines of a parameterized mathematical model. To the extent that only a small subset of real-world environments correspond to the abstract environments in economic models, it is not clear whether the generality of such abstract models is a virtue. Valid generalizations for the set of all real-world environments may not hold more generally within the larger set of fictitious environments parameterized in the economic model, and would therefore be impossible to discover in an unrestricted model space. These potential disadvantages are counterbalanced by one clear benefit of formal modeling, which is the opportunity to observe and interpret functional links between environments and the relative success of different decision procedures. In spite of methodological differences, both simulation and formal modeling converge around the proposition of ecological rationality as an organizing framework for normative decision theory and its application.

Finally, we note that the model of adaptive ignoring without cognitive limitations presented in this paper is both distinct from, and complementary to, bounded rationality explanations for behavioral insensitivity to relevant information. The focus of this investigation was squarely on the role of the external environment in favoring different amounts of usage of available information. Joint explanations concerning the match between external stochastic payoff structure and internal cognitive processes provide perhaps the most promising direction to account for the demonstrated real-world success of simple decision algorithms.

References

Berg, N. (2005). Decision-making environments in which unboundedly rational decision makers choose to ignore relevant information. *Global Business and Economics Review, 7*(1), 59–73.

Berg, N., & Hoffrage, U. (2008). Rational ignoring with unbounded cognitive capacity. *Journal of Economic Psychology, 29,* 792–809.

Bookstaber, R., & Langsam, J. (1985). On the optimality of coarse behavior rules. *Journal of Theoretical Biology, 116*, 161–193.

Brooks, R. A. (1991). New approaches to robotics. *Science, 253*, 1227–1232.

Bullock, S., & Todd, P. M. (1999). Made to measure: Ecological rationality in structured environments. *Minds and Machines, 9*(4), 497–541.

Clarke, D. D., Ward, P., & Truman, W. (2002). In-depth accident causation study of young drivers, Report no. TRL542, TRL Limited, Crowthorne, Berkshire.

Connolly, T. (1999). Action as a fast and frugal heuristic. *Minds and Machines, 9*(4), 479–496.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In: G. Gigerenzer, P. M. Todd & the ABC Research Group, *Simple heuristics that make us smart* (pp. 97–118). New York: Oxford University Press.

Dobbie, K. (2002). Fatigue-related crashes: An analysis of fatigue-related crashes on Australian roads using an operational definition of fatigue, Australian transport safety bureau report no. OR23.

Forster, M. R. (1999). How do simple rules 'fit to reality' in a complex world? *Minds and Machines, 9*(4), 543–564.

Gigerenzer, G. & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science, 1*, 107–143

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650–669.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.

Gigerenzer, G., & Selten, R. (Eds.) (2001). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109*, 75–90.

Hansen, M. H., & Bin, Y. (1996). Model selection and minimum description. *Journal of the American Statistical Association, 96*, 746–774.

Hogarth, R. M. (2001). *Educating instituition*. Chicago: University of Chicago Press.

Hogarth, R. M., & Karelaia, N. (2005). Ignoring information in binary choice with continuous variables: When is less "more"? *Journal of Mathematical Psychology, 49*, 115–124.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Luce, D., & Raiffa, H. (1957). *Games and decisions*. New York: Wiley.

Mahowald, M. W. (2000). Eyes wide shut: The dangers of sleepy driving. *Minnesota Medicine, 83*, 25–30.

Martignon, L., & Hoffrage, U. (2002). Fast, frugal and fit: Simple heuristics for paired comparison. *Theory and Decision, 52*, 29–71.

Martignon, L., Katsikopoulos, K. V., & Woike, J. K. (2008). Categorization with limited resources: A family of simple heuristics. *Journal of Mathematical Psychology, 52*(6), 352–361.

Martignon, L., & Schmitt, M. (1999). Simplicity and robustness of fast and frugal heuristics. *Minds and Machines, 9*(4), 565–593.

Mohan, D. (2001). Traffic safety and thirty years of biomechanics research: A personal adventure. In *International Research Council on the Biomechanics of Impact Proceedings* (pp. 1–12).

Pack, A. I., Pack, A. M., Rodgman, E., Cucchiara, A., Dinges, D. F., & Schwab, C. W. (1995). Characteristics of crashes attributed to the driver having fallen asleep. *Accident Analysis and Prevention, 27*(6), 769–775.

Raiffa, H. (1986). *Decision analysis: Introductory lectures on choices under uncertainty*. New York: Random House.

Raiffa, H., & Shlaifer, R. (1961). *Applied statistical decision theory*. Cambridge, MA: Harvard University Press.

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review, 112*, 610–628.

Simon, H. A. (1982). *Models of bounded rationality*. Cambridge: MIT Press.

Todd, P. M. (1999). Simple inference heuristics versus complex decision machines. *Minds and Machines, 9*(4), 461–477.

Todd, P. M. (2000) The ecological rationality of mechanisms evolved to make up minds. *American Behavioral Scientist, 43*, 940–956.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In: D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153–160). New York: Cambridge University Press.

Zellner, A., Keuzenkamp, H., & McAleer, M. (Eds.). (2001). *Simplicity, inference and modelling: Keeping it sophisticatedly simple*. Cambridge: Cambridge University Press.