ECONOMIC
JOURNAL

The Economic Journal, **125** (*February*), 184–202. Doi: 10.1111/ecoj.12130

A TESTABLE THEORY OF IMPERFECT PERCEPTION*

Andrew Caplin and Daniel Martin

We provide a characterisation of choice behaviour generated by a Bayesian expected utility maximiser. The observable signature of this standard model is the impossibility of raising utility by switching wholesale from one action to another. We provide applications to robustness, to the recovery of utility from choice data and to model classification.

There is a standard approach to modelling signal processing and choice. Decision-makers start out with prior beliefs concerning an underlying state of the world that determines the pay-offs to all actions. They receive additional signals concerning this state and update their priors in a Bayesian manner. Their final choice of action maximises expected utility given these posterior beliefs.

From an applied point of view, the devil is in the details. In a typical application it is impossible to know what form private signals take, let alone how well they are understood. By way of example, consider jurors in a trial. Even if one tightly monitors the trial, one cannot know how the proceedings were translated into informative signals concerning the guilt or innocence of the defendant. Nor can one know how such signals were processed. Is it accurate to model jurors as having perceived all information perfectly? Should we instead model them as perceiving noisy signals? If so, what form should these signals take and from what distribution should they be drawn? When there is no clear answer to these questions, it seems sensible to ask what we can say if we make no assumptions about the exact form of information processing.

In this study, we identify the limits that the standard model of Bayesian expected utility (BEU) maximisation places on behaviour if no assumptions are made about underlying signals. These limits are summarised by a set of linear inequalities on the state-dependent stochastic choice of actions. These 'no improving action switches' (NIAS) inequalities have a simple interpretation: it is impossible to improve utility by making wholesale switches from one action to another. While the necessity of the NIAS inequalities is clear, the sufficiency of the inequalities is more surprising. If they apply, there is always some specification of private signals and utilities that rationalises the data.

We present three applications of the NIAS inequalities. In the first application, we make predictions for behaviour that are 'robust' to the exact form of signal processing.

* Corresponding author: Daniel Martin, Paris School of Economics, 48 Blvd. Jourdan, 75014 Paris, France. Email: daniel@martinonline.org.

We thank Dirk Bergemann, Colin Camerer, David Cesarini, Olivier Compte, Mark Dean, Sen Geng, Paul Glimcher, Philippe Jehiel, Paola Manzini, Marco Mariotti, Stephen Morris, Antonio Rangel, Collin Raymond, Michael Richter, Natalia Shestakova, Jonathan Weinstein, the co-editor and two anonymous referees for valuable comments.

This is analogous to the robust predictions for games of incomplete information found in Bergemann and Morris (2013*a*) (see subsection 6.1). In a jury voting example, we show that when the defendant is more likely to be innocent, the standard model robustly predicts low levels of Type I error (voting to convict an innocent person) but puts no restriction on the degree of Type II error (voting to acquit a guilty person).

In our second application we illustrate the bounds that the NIAS inequalities place on utility. Even when ordinal rankings cannot be identified, the NIAS inequalities establish bounds on the relative strength of preference for one prize over another.

In our final application, we use the NIAS inequalities to identify whether or not prominent forms of boundedly rational behaviour can be rationalised with a standard Bayesian model. This is a non-trivial question as there are many different models of choice that produce the same behaviour (Richter, 2011).[1] We show that models based on 'consideration sets' (such as those of Masatlioglu *et al.*, 2012; Manzini and Mariotti, 2014) produce behaviours that violate the NIAS inequalities, hence cannot be so rationalised. The same applies to behaviours generated by procedural models of list order search, such as that of Rubinstein and Salant (2006). We show also that the standard logit model of discrete choice, although sometimes motivated by cognitive limits (McKelvey and Palfrey, 1995), is inconsistent with the standard Bayesian model.

The second and third applications are similar in spirit to the work of Salant and Rubinstein (2008), who show conditions under which choices that are distorted by a 'frame' can be explained by a transitive binary relation, and Rubinstein and Salant (2011), who present a more general framework for determining a welfare ordering from behavioural data sets. Unlike this work, our article is based on the study of unobservable information that is 'relevant to the rational assessment of the alternatives and thus should not be regarded as a frame' and thus lies 'outside the scope' of their work (Salant and Rubinstein, 2008, p. 1288).[2] Our technical results are also quite different, as we model the stochastic choice of actions, not deterministic preference orderings.

In Section 1 we introduce our formal model and establish that the NIAS inequalities characterise BEU maximisation. In Section 2 we demonstrate use of these inequalities in restricting behaviour in a jury trial example. In Section 3 we show how to use these inequalities to bound utilities using choice data. In Section 4 we use the inequalities to classify behaviour according to whether or not they are rationalisable by the standard Bayesian model.

In the main result, we treat prior probabilities as observable and known to the decision-maker. In Section 5 we extend the characterisation to cases in which prior probabilities are unobservable or decision-makers have a subjective prior. We discuss strategic analogues and additional applications in Section 6.

[1] To illustrate that this is not always obvious, note that a simple satisficing model produces the same behaviour as a standard model of maximisation in which 'satisfactory' items are mutually indifferent and superior to unsatisfactory items (Tyson, 2008).

[2] Caplin and Martin (2013*b*) combine these traditions by considering the case where frames can distort attention to decision relevant information.

1. BEU Representations

1.1. *Decision Problem and Data Set*

A decision problem is defined by a quadruple (A, X, Ω, μ), where A is a finite set of actions, X is a finite set of prizes, Ω is a finite set of states, and μ is the probability distribution (prior) over states.

Actions may take many different forms:

(*i*) voting guilty or not guilty in a trial;
(*ii*) buying or not buying a good;
(*iii*) choosing one of A positions in a list (Rubinstein and Salant, 2006; Caplin *et al.*, 2011; Caplin and Martin, 2013*a,b*); and
(*iv*) choosing one of A possible prices (Martin, 2012) or making one of available guesses concerning the number of blue balls in a display (Caplin and Dean, 2013, 2014).

The underlying state of the world $\omega \in \Omega$ specifies the precise connection between actions and prizes. We define each state as a function $\omega : A \rightarrow X$. The prior probabilities are summarised by $\mu \in \Gamma = \Delta(\Omega)$, with μ_ω the prior probability of state ω.

The data set is a joint distribution over states and action choices. Technically speaking, given (A, X, Ω), a state-dependent stochastic data set q identifies the probability distribution over action choices as it depends on the state,

$$q : \Omega \rightarrow \Delta(A).$$

Let q_ω^a be the probability of action a when the state is ω.

1.2. *Observability*

In the first four sections of this study, we assume that the outside observer (econometrician or model-builder) can observe both the decision problem and the data set. Because the states in which choices are made is observable, the outside observer can drop impossible states, so that $\mu_\omega > 0$ for all $\omega \in \Omega$. In certain applied settings, the prior probabilities can be estimated, induced or elicited, and it is this case that we consider in the main body of the text. In Section 5 we consider cases in which the prior probabilities are unobservable or decision-makers have a subjective prior.

The applications found in this article are theoretical in nature and do not require collecting state-dependent stochastic choice data. The feasibility of collecting such data is established in the experiments of Martin (2012), Caplin and Dean (2013, 2014) and Caplin and Martin (2013*a,b*). With additional assumptions concerning stationarity over time and homogeneity across individuals, such data can be constructed also from standard observational data.

1.3. *Unobservables*

We ask whether unobservables exist that explain the observables as if the decision-maker:

(*i*) knows the decision problem (including the prior probabilities);
(*ii*) obtains signals (private information) about the state;
(*iii*) uses Bayes' rule to update the prior; and
(*iv*) maximises expected utility based on the updated beliefs.

In technical terms, there are three unobservables: an expected utility function; a perception function and a choice function.[3]

The expected utility function is defined in standard manner on the prize space, $U : X \to \mathbb{R}$. The utility that results from taking action a in state ω can then be generated by taking the composition of the utility function and the map between actions and prizes in this state,

$$U^a_\omega = U[\omega(a)].$$

The fact that the utility function is defined over prizes implies that if the prize that action a yields in state ω is the same as the prize that action b yields in state v, then $U^a_\omega = U^b_v$. While Theorem 1 does not require this assumption, in applications it provides much of the bite of the NIAS inequalities.

The second and less standard unobservable is a perception function that produces posterior beliefs. The goal of our theory is not to specify the exact process of signal extraction and signal processing but rather to characterise general properties associated with the standard model of signal processing and choice. To that end, we cut through the details of the signal extraction and processing technology and work directly with distributions over posterior beliefs.[4] Formally, we define a perception function as a mapping from states of the world into $\Delta(\Gamma)$, the probability distributions over the set of beliefs Γ with finite support,

$$\pi : \Omega \to \Delta(\Gamma).$$

A generic posterior belief is given by $\gamma \in \Gamma$, with γ_ω the posterior probability of state ω. Letting $\pi_\omega(\gamma)$ be the probability of posterior γ in state ω, we define $\Gamma(\pi)$ as the set of possible posteriors for perception function π,

$$\Gamma(\pi) = \cup_{\omega\in\Omega}\{\gamma \in \Gamma | \pi_\omega(\gamma) > 0\}.$$

The final unobservable is a choice function C that maps possible posteriors to action probabilities,

$$C : \Gamma(\pi) \to \Delta(A).$$

We let $C^a(\gamma)$ denote the probability of choosing action a with posterior $\gamma \in \Gamma(\pi)$.

[3] As noted above, in Section 5 we add the prior as a fourth unobservable but for now we treat $\mu \in \Gamma$ as observed.

[4] Kamenica and Gentzkow (2011) show that it is without loss of generality to work directly with posterior beliefs rather than with underlying signals.

1.4. *BEU Representation*

For π and C to provide a possible explanation of the data set requires that their composition generates the data set. To make them consistent with the standard model of signal processing, we insist that π satisfies Bayes's rule. To ensure consistency with the standard model of choice, we insist that C maximises expected utility. A BEU maximising representation is defined by satisfying these three conditions.

DEFINITION 1. *(π, C, U) is a BEU representation of (A, X, Ω, μ, q) if it satisfies:*

(*i*) *Data matching: For all $\omega \in \Omega$ and $a \in A$,*

$$q_{\omega}^{a} = \sum_{\gamma \in \Gamma(\pi)} \pi_{\omega}(\gamma) C^{a}(\gamma).$$

(*ii*) *Bayesian updating: For all $\omega \in \Omega$ and $\gamma \in \Gamma(\pi)$,*

$$\gamma_{\omega} = \frac{\mu_{\omega} \pi_{\omega}(\gamma)}{\sum_{\upsilon \in \Omega} \mu_{\upsilon} \pi_{\upsilon}(\gamma)}.$$

(*iii*) *Maximisation: For all $\gamma \in \Gamma(\pi)$ and $a \in A$ such that $C^{a}(\gamma) > 0$,*

$$\sum_{\omega \in \Omega} \gamma_{\omega} U_{\omega}^{a} \geq \sum_{\omega \in \Omega} \gamma_{\omega} U_{\omega}^{b} \text{ all } b \in A,$$

with the inequality strict for some $\gamma \in \Gamma(\pi)$ and $a, b \in A$.

We require strictness in the utility comparison of some pair of actions in some state to prevent the conditions from being trivially satisfied by a utility function with all prizes indifferent.

1.5. *The NIAS Theorem*

In our data set, BEU maximisation is characterised by the impossibility of raising utility by switching wholesale from one action to another. This condition is formalised in the NIAS inequalities.

DEFINITION 2. *Utility function $U : X \to \mathbb{R}$ satisfies the NIAS inequalities with respect to (A, X, Ω, μ, q), if,*

$$\sum_{\omega \in \Omega} \mu_{\omega} q_{\omega}^{a} U_{\omega}^{a} \geq \sum_{\omega \in \Omega} \mu_{\omega} q_{\omega}^{a} U_{\omega}^{b},$$

for all $a, b \in A$, and the inequality is strict for some $a, b \in A$.

The main result is that the NIAS inequalities characterise BEU representations. The proof of the necessity of the NIAS inequalities follows directly from the definition of a BEU representation. The inequalities in maximisation become the inequalities in NIAS when we use data matching to turn the unobservable π and C into the observable q.

The sufficiency of the NIAS inequalities is established by constructing a BEU representation based on any utility function that satisfies the NIAS inequalities. The first step is to construct a perception function using the observed data. In this step, we use the data to construct a 'revealed' posterior belief for each action. The second step is to determine a choice function that rationalises the data on the basis of this perception function. This step is straightforward when all actions are associated with distinct posteriors. However, we must also allow for cases in which the perception function maps more than one action to the same posterior. It is to cover this case that we need the choice function C to allow for mixing.

THEOREM 1. *(A, X, Ω, μ, q) has a BEU representation if and only if there exists $U : X \to \mathbb{R}$ satisfying the NIAS inequalities.*

Proof. Necessity: Suppose that (π, C, U) define a BEU of (A, X, Ω, μ, q). We show directly that $U : X \to \mathbb{R}$ must satisfy the NIAS inequalities. Note first that from maximisation, given any $\gamma \in \Gamma(\pi)$ and $a \in A$,

$$C^a(\gamma)\left(\sum_{\omega\in\Omega}\gamma_\omega U^a_\omega\right) \geq C^a(\gamma)\left(\sum_{\omega\in\Omega}\gamma_\omega U^b_\omega\right) \text{ all } b \in A.$$

Adding up across $\gamma \in \Gamma(\pi)$, using the Bayesian updating property to substitute for γ_ω, changing order of addition, and cancelling common terms $\sum_{\upsilon\in\Omega}\mu_\upsilon\pi_\upsilon(\gamma) > 0$ in all denominators, we derive,

$$\sum_{\omega\in\Omega}\mu_\omega\left[\sum_{\gamma\in\Gamma(\pi)}\pi_\omega(\gamma)C^a(\gamma)\right]U^a_\omega \geq \sum_{\omega\in\Omega}\mu_\omega\left[\sum_{\gamma\in\Gamma(\pi)}\pi_\omega(\gamma)C^a(\gamma)\right]U^b_\omega \text{ all } b \in A.$$

We now use data matching to substitute for the inner summations and derive,

$$\sum_{\omega\in\Omega}\mu_\omega q^a_\omega U^a_\omega \geq \sum_{\omega\in\Omega}\mu_\omega q^a_\omega U^b_\omega,$$

verifying the all NIAS inequalities hold at least weakly. To confirm that at least one such inequality is strict, pick $a, b \in A$ and $\gamma \in \Gamma(\pi)$ with $C^a(\gamma) > 0$ for which maximisation holds strictly,

$$C^a(\gamma)\left(\sum_{\omega\in\Omega}\gamma_\omega U^a_\omega\right) > C^a(\gamma)\left(\sum_{\omega\in\Omega}\gamma_\omega U^b_\omega\right).$$

Repeating other steps from this point forward reveals that the corresponding NIAS inequality holds strictly,

$$\sum_{\omega\in\Omega}\mu_\omega q^a_\omega U^a_\omega > \sum_{\omega\in\Omega}\mu_\omega q^a_\omega U^b_\omega.$$

Sufficiency: Consider a function $U : X \to \mathbb{R}$ that satisfies the NIAS inequalities with respect to (A, X, Ω, μ, q). We now identify $\bar{\pi}$ and $\bar{C}$ such that $(\bar{\pi}, \bar{C}, U)$ provides a BEU representation of (A, X, Ω, μ, q). First, define chosen actions $\bar{A} \subset A$ as all those for which $q^a_\omega > 0$ some $\omega \in \Omega$. Given $a \in \bar{A}$ and $\omega \in \Omega$, define the corresponding posterior $\bar{\gamma}^a_\omega$ by,

$$\bar{\gamma}^a_\omega \equiv \frac{\mu_\omega q^a_\omega}{\sum_{\upsilon\in\Omega}\mu_\upsilon q^a_\upsilon}.$$

To complete the construction, first partition the set of possible actions into $P \leq J$ sets $\bar{A}(p)$ with identical posteriors $\bar{\gamma}(p)$ within each such set and distinct posteriors in each such set. Hence, $a,\ b \in \bar{A}(p)$ if and only if $\bar{\gamma}^a = \bar{\gamma}^b = \bar{\gamma}(p)$, so for $a,\ b \in \bar{A}(p)$ and $\omega \in \Omega$,

$$\sum_{\upsilon\in\Omega}\mu_\upsilon q^b_\upsilon = \mu_\omega q^b_\omega\left(\frac{\sum_{\upsilon\in\Omega}\mu_\upsilon q^a_\upsilon}{\mu_\omega q^a_\omega}\right). \tag{1}$$

Now define the perception function to have domain $\Gamma(\bar{\pi}) = \cup_{p=1}^{P}\bar{\gamma}(p)$, and specific values,

$$\bar{\pi}_\omega[\bar{\gamma}(p)] = \sum_{b\in\bar{A}(p)} q^b_\omega. \tag{2}$$

Finally, define the choice function to satisfy,

$$\bar{C}^a[\bar{\gamma}(p)] = \begin{cases} \dfrac{\sum_{\upsilon\in\Omega}\mu_\upsilon q^a_\upsilon}{\sum_{b\in\bar{A}(p)}\sum_{\upsilon\in\Omega}\mu_\upsilon q^b_\upsilon} \in (0,1] & \text{if } a \in \bar{A}(p); \\ 0 & \text{if } a \notin \bar{A}(p). \end{cases} \tag{3}$$

To confirm that this construction identifies a BEU, we first establish data matching. By construction, given $\omega \in \Omega$ and $a \in \bar{A}(p)$, we know that $\bar{C}^a[\bar{\gamma}(p)] = 0$ unless $a \in \bar{A}(p)$. Hence, for each $\omega \in \Omega$ and $a \in \bar{A}(p)$,

$$\sum_{\gamma\in\Gamma(\bar{\pi})}\bar{\pi}_\omega(\gamma)\bar{C}^a(\gamma) = \sum_{p=1}^{P}\bar{\pi}_\omega[\bar{\gamma}(p)]\bar{C}^a[\bar{\gamma}(p)] = \sum_{p=1}^{P}\left[\sum_{b\in\bar{A}(p)} q^b_\omega\right]\left[\frac{\sum_{\upsilon\in\Omega}\mu_\upsilon q^a_\upsilon}{\sum_{b\in\bar{A}(p)}\sum_{\upsilon\in\Omega}\mu_\upsilon q^b_\upsilon}\right], \tag{4}$$

which follows directly from substitution of (2) and (3). Now note from (1) that

$$\sum_{b\in\bar{A}(p)}\sum_{\upsilon\in\Omega}\mu_\upsilon q^b_\upsilon = \sum_{b\in\bar{A}(p)}\mu_\omega q^b_\omega\left(\frac{\sum_{\upsilon\in\Omega}\mu_\upsilon q^a_\upsilon}{\mu_\omega q^a_\omega}\right). \tag{5}$$

Substitution of (5) in the denominator in (4) yields

$$\begin{aligned}\sum_{p=1}^{P}\bar{\pi}_\omega[\bar{\gamma}(p)]\bar{C}^a[\bar{\gamma}(p)] &= \sum_{p=1}^{P}\left[\sum_{b\in\bar{A}(p)} q^b_\omega\right]\left[\frac{\mu_\omega q^a_\omega}{\sum_{b\in\bar{A}(p)}\mu_\omega q^b_\omega}\right] = \sum_{p=1}^{P}\frac{\sum_{b\in\bar{A}(p)}\mu_\omega q^b_\omega q^a_\omega}{\sum_{b\in\bar{A}(p)}\mu_\omega q^b_\omega} \\ &= q^a_\omega\sum_{p=1}^{P}\left[\frac{\sum_{b\in\bar{A}(p)}\mu_\omega q^b_\omega}{\sum_{b\in\bar{A}(p)}\mu_\omega q^b_\omega}\right] = q^a_\omega,\end{aligned}$$

in confirmation of data matching.

To confirm Bayesian updating, note as a result of data matching that for all $\omega \in \Omega$, $1 \leq p \leq P, \bar{\gamma}(p) \in \Gamma(\pi)$ and $a \in \bar{A}(p)$,

$$\bar{\gamma}_\omega(p) = \frac{\mu_\omega q_\omega^a}{\sum_{\upsilon\in\Omega}\mu_\upsilon q_\upsilon^a} = \frac{\mu_\omega\bar{\pi}_\omega[\bar{\gamma}(p)]\bar{C}^a[\bar{\gamma}(p)]}{\sum_{\upsilon\in\Omega}\mu_\upsilon\bar{\pi}_\upsilon[\bar{\gamma}(p)]\bar{C}^a[\bar{\gamma}(p)]} = \frac{\mu_\omega\bar{\pi}_\omega[\bar{\gamma}(p)]}{\sum_{\upsilon\in\Omega}\mu_\upsilon\bar{\pi}_\upsilon[\bar{\gamma}(p)]}.$$

Finally, note that for each $\omega \in \Omega, 1 \le p \le P$ and $a \in \bar{A}(p)$

$$\mu_\omega q_\omega^a = \bar{\gamma}_\omega(p)\sum_{\upsilon\in\Omega}\mu_\upsilon q_\upsilon^a.$$

Substitution in the NIAS inequalities and division by the constant $\sum_{\upsilon\in\Omega}\mu_\upsilon q_\upsilon^a > 0$ yield

$$\sum_{\omega\in\Omega}\bar{\gamma}_\omega(p)U_\omega^a \ge \sum_{\omega\in\Omega}\bar{\gamma}_\omega(p)U_\omega^b,$$

for all $\omega \in \Omega$, $1 \le p \le P$ and $a \in \bar{A}(p)$, with the inequality strict for some a, $b \in \bar{A}$. This establishes maximisation and completes the proof.

The above result is robust in a number of senses. First, in typical cases the perception function defined in (2) above will assign just one posterior to each action. In such cases the choice function specified in the proof is deterministic, so mixing is not required. Second, to strengthen the requirements for a BEU representation to ensure that all actions are uniquely optimal at the corresponding posteriors requires only the corresponding strengthening of the NIAS inequalities. Finally, while we treat the prior as observable in the above proof, we show in Section 5 that an analogous result applies when the prior is not observable.

Note that each NIAS inequality imposes a linear constraint on prize utilities. Existence of a utility function that validates all such inequalities corresponds to establishing non-emptiness of the intersection of $(J-1)^2$ linear inequalities. This can be checked using standard linear programming methods.

Afriat (1967) similarly provided a set of data-defined linear inequalities such that a solution to the inequalities exists if and only if a non-satiated utility function exists that rationalises the data. While not directly comparable because they are based on deterministic choices from budget sets, Afriat's inequalities have a conceptual link with our constraints in that they both treat the utility function as unobservable.

2. Robust Prediction

Our first application of the NIAS inequalities is to produce predictions for behaviour that are robust to assumptions about information processing. To illustrate, we use an example of a juror in a criminal trial, as inspired by the lead example in Kamenica and Gentzkow (2011) and Bergemann and Morris (2013*a,b*). A defendant is either innocent (state *I*) or guilty (state *G*). The juror can acquit (action *A*) or convict (action *C*). The prior probability of innocence is $\mu_I \in (0, 1)$. Finally, two parameters define the data set: the probability of voting to acquit when the defendant is either innocent or guilty, $\alpha_I \equiv q_I^A$ and $\alpha_G \equiv q_G^A$.

In analysing the outcome of the trial, the following statistics are of central interest:

(*i*) the overall probability of voting to acquit is $\mu_I\alpha_I + (1-\mu_I)\alpha_G$;

(*ii*) the probability of voting to convict an innocent party, known as Type I error, is $1 - \alpha_I$;
(*iii*) the probability of voting to acquit a guilty party, known as Type II error, is α_G; and
(*iv*) the overall probability of a mistaken vote is $\mu_I(1-\alpha_I) + (1-\mu_I)\alpha_G$.

In this illustration, we use the NIAS inequalities to indicate when it is possible to robustly predict a low level of Type I error (voting to convict an innocent party) or Type II error (voting to acquit a guilty party).

2.1. *NIAS Inequalities and Limits on Behaviour*

We assume that the juror prefers voting correctly: to acquit if the defendant is innocent and to convict if the defendant is guilty. As a starting point, we follow Kamenica and Gentzkow (2011) and Bergemann and Morris (2013*a,b*) in assuming that the juror only cares about voting correctly and in normalising the utility of voting correctly to 1 and the utility of voting incorrectly to 0.

It is intuitively clear that the choice data produced by a BEU maximiser must satisfy certain conditions. For example, if the probability of innocence is high, there must be high likelihood of voting to acquit. The NIAS inequalities provide these intuitive restrictions in a concise and precise manner. The NIAS inequality for voting to acquit simplifies to

$$\alpha_I \geq \left(\frac{1-\mu_I}{\mu_I}\right)\alpha_G.$$

The NIAS inequality for voting to convict simplifies to

$$\alpha_I \geq \left(2 - \frac{1}{\mu_I}\right) + \left(\frac{1-\mu_I}{\mu_I}\right)\alpha_G.$$

As always, at least one inequality must be strict.[5]

With an even prior ($\mu_I = 0.5$), the NIAS inequalities assert that $\alpha_I > \alpha_G$, which means that the juror must be correct strictly more than 50% of the time. It is clear that this is necessary for there to exist a BEU representation. That it is also sufficient shows that being very good at correctly identifying an innocent party can be consistent with being very bad at correctly identifying a guilty party (and *vice versa*) without requiring non-Bayesian reasoning or caring about the type of error committed. If correct on innocence 99% of the time ($\alpha_I = 0.99$), such a juror can be incorrect on guilt up to 98% of the time ($\alpha_G = 0.98$).

Despite the reach of the standard model, precisely 50% of all conceivable data sets are ruled out by this condition, as illustrated by the shaded region beneath the main diagonal in Figure 1. As the prior becomes more uneven, the implied restrictions on choice data become stronger. To illustrate, the larger shaded region in Figure 1

[5] With $\mu_I < 0.5$ the second inequality is non-binding and must hold strictly, whereas with $\mu_I > 0.5$ the first inequality is non-binding, hence strict. When $\mu_I = 0.5$, the inequalities are identical so that both must hold strictly.

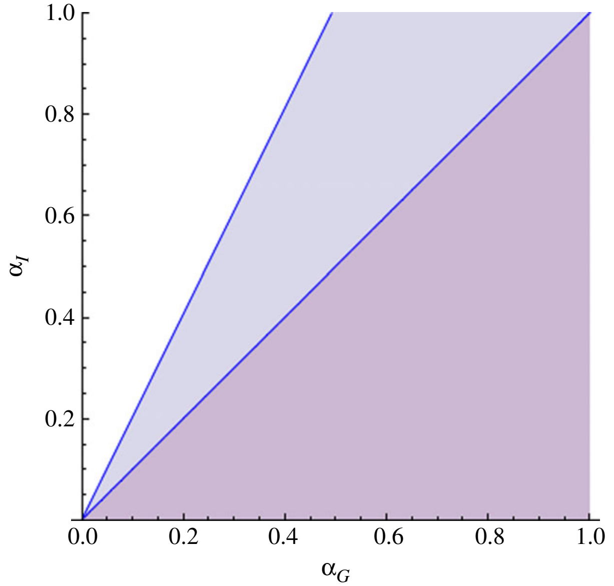

Fig. 1. *Robust Predictions for* $\mu_I = 0.5$ and $\mu_I = 1/3$.

indicates the choice probabilities that cannot be rationalised when the prior probability of innocence is $\mu_I = 1/3$, in which case the relevant inequality identifying existence of a BEU is

$$\alpha_I \geq 2\alpha_G.$$

Note that 75% of all conceivable data sets are ruled out by this condition.

2.2. *Type I and Type II Error*

Note that when $\mu_I = 1/3$, there is no absolute restriction on Type I error (voting to convict an innocent person), whereas the rate of Type II error (voting to acquit a guilty person) cannot be above 50%. As the probability of innocence falls, the upper bound on Type II error falls.

The bounds on errors are interdependent. If $\alpha_I = 0$, so that the juror always makes a Type I error, then must be no Type II error. If $\alpha_I = 1$, so that the juror never makes Type I error, then there can be up to a 50% rate of Type II error. Between these extremes, the relationship between the Type I error and the maximum Type II error is linear.

Finally, note that when the defendant is more likely to be innocent, say $\mu_I = 2/3$, the relationship between Type I and Type II errors reverses. For example, there is no absolute restriction on Type II errors, whereas the Type I error rate cannot be above 50%.

2.3. *Caring About Errors*

We now allow for the juror to dislike Type I errors (voting to convict an innocent defendant) differently from Type II errors (voting to acquit a guilty defendant).

We partially continue the normalisation from before, so that the utility of voting correctly is 1 and the utility of making a Type II error is 0. However, now the utility of making a Type I error is $u < 1$. For values of $u \in (0, 1)$, the juror dislikes Type II error more than Type I error, while the converse holds for $u < 0$.

The NIAS inequality for voting to acquit now simplifies to,

$$\alpha_I \geq \left[\frac{1-\mu_I}{(1-u)\mu_I}\right]\alpha_G,$$

while the NIAS inequality for voting to convict simplifies to

$$\alpha_I \geq \left[\frac{2-u}{1-u} - \frac{1}{(1-u)\mu_I}\right] + \left[\frac{1-\mu_I}{(1-u)\mu_I}\right]\alpha_G.$$

The restrictiveness of the NIAS inequalities now depends both on the prior and on the value of u. With $\mu_I = 0.5$ and $u = -1$, the inequality is,

$$\alpha_I \geq \frac{1}{2} + \frac{1}{2}\alpha_G.$$

This implies that the juror must correctly vote to acquit the innocent at least 50% of the time, with this bound becoming ever more restrictive the more they incorrectly acquit the guilty. Keeping this same prior and increasing the asymmetry in the utility function to the point where $u = -9$, the constraint becomes,

$$\alpha_I \geq \frac{9}{10} + \frac{1}{10}\alpha_G,$$

so that the juror must correctly acquit the innocent at least 90% of the time. Fully 95% of all conceivable data sets are ruled out by this condition, as illustrated by the larger shaded region in Figure 2.

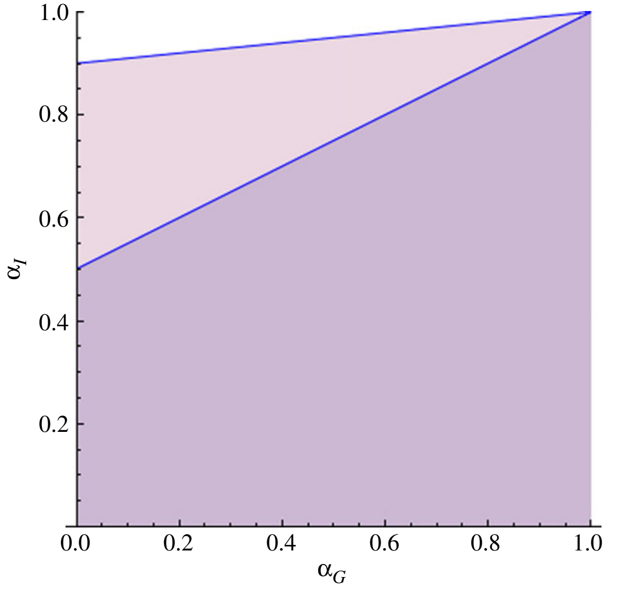

Fig. 2. *Robust Predictions for* $u = -1$ and $u = -9$, when $\mu_I = 0.5$.

Looking across these examples, with $\mu_I = 0.5$ and $u = 0$, there are no restrictions on Type I error but with, $u = -1$, we see a substantial restriction on the rate of Type I error. As would be expected, with $u = -9$, we see even tighter restrictions on the probability of a Type I error.

3. Bounds on Expected Utility

The NIAS inequalities allow us to use choice data to place bounds on unknown utilities even when the exact form of information processing is unspecified. Even when ordinal rankings cannot be determined, the NIAS inequalities establish bounds on the relative strength of preference for one prize over another.

To illustrate, consider again the example of the previous Section, but where the relative importance of Type I and Type II errors in the juror's utility function is unknown. Again this is determined by a single parameter, $u < 1$, the utility of Type I error.

In technical terms, we use the NIAS inequalities to constrain u. The inequality that makes voting to acquit at least as beneficial to the juror as voting to convict simplifies to,

$$u \leq 1 - \left(\frac{1-\mu_I}{\mu_I}\right)\left(\frac{\alpha_G}{\alpha_I}\right).$$

The NIAS inequality that makes voting to convict at least as beneficial as voting to acquit simplifies to,

$$u \geq 1 - \left(\frac{1-\mu_I}{\mu_I}\right)\left(\frac{1-\alpha_G}{1-\alpha_I}\right).$$

Overall the requirement is,

$$1 - \left(\frac{1-\mu_I}{\mu_I}\right)\left(\frac{\alpha_G}{\alpha_I}\right) \geq u \geq 1 - \left(\frac{1-\mu_I}{\mu_I}\right)\left(\frac{1-\alpha_G}{1-\alpha_I}\right),$$

with at least one inequality strict.

To see how the NIAS inequalities constrain relative error costs, consider first a case with $\mu_I = 0.5$, $\alpha_I = 2/3$ and $\alpha_G = 1/3$. In this case the inequalities assert

$$\frac{1}{2} \geq u \geq -1.$$

These constraints do not pin down whether Type I or Type II error is worse because the utility of Type II error is zero. Rather, the inequalities constrain the ratio of the losses associated with Type I errors relative to Type II errors.

In fact, with $\mu_I = 0.5$, the bounds are on the opposite sides of zero for any choice data,

$$1 - \frac{\alpha_G}{\alpha_I} \geq u \geq 1 - \frac{1-\alpha_G}{1-\alpha_I}.$$

Hence, one cannot know ordinal rankings of Type I and Type II error for this prior. However, in the limit as α_I approaches α_G, the utility function is almost exactly pinned down. For $\mu_I = 0.5$, in the limit as α_I approaches α_G, the juror must be close to indifferent between Type I errors and Type II errors. The bounds tighten when the conditional choice probabilities get closer because there is less room for imperfect perception to explain variation in choice.

When the prior is uneven, it may be possible to pin down which type of error is worse and also to provide bounds on the extent of the difference. For example, with $\mu_I = 1/4$, the inequality becomes,

$$1 - 3\left(\frac{\alpha_G}{\alpha_I}\right) \geq u \geq 1 - 3\left(\frac{1 - \alpha_G}{1 - \alpha_I}\right).$$

For example, when $\alpha_I = 2/3$ and $\alpha_G = 1/3$, the requirement is,

$$-\frac{1}{2} \geq u \geq -5.$$

In this case it is known for sure that Type I errors are regarded as worse than Type II errors. Again, the precise extent of this preference is not known but, in the limit, as α_I falls towards α_G, the utility function is almost fully pinned down. In this case, the utility bounds tighten around $u = -2$.

4. Model Classification

The NIAS inequalities can be used to classify models of choice as either consistent or inconsistent with BEU maximisation directly from the choice data they produce. To match a typical application in the literature on bounded rationality, we now consider a consumer choosing between two goods x_1 and x_2. The consumer strictly prefers product x_1, and the corresponding utility function is normalised to $U(x_1) = 1$ and $U(x_2) = 0$. However, the goods look somewhat similar and are put side-by-side on a shelf. Hence, it may be hard for the consumer to determine which good is which.

The consumer can choose the good on the left (action L) or choose the good on the right (action R). The preferred good is either on the left (state l) or on the right (state r), with the prior $\mu_l \in (0, 1)$ identifying the probability that it is on the left. The two parameters that define the data set are $\lambda_l \equiv q_l^L$, $\lambda_r \equiv q_r^L$, the probabilities of picking the good on the left in either state.

4.1. *Stochastic Consideration of Prizes*

The bounded rationality literature offers many approaches to modelling imperfect perception. Manzini and Mariotti (2014) propose a form of stochastic consideration where a prize x_n is considered with probability $\eta_n \in (0, 1)$ and the optimal option inside the consideration set is chosen.[6] The default choice for an empty consideration set is left unspecified. In this example, we assume that the default choice gives the decision-maker an inferior good x_3 with certainty.

The data produced by this theory reflect the fact that x_1 is chosen if considered and x_2 is chosen only if it is the only prize considered,

[6] Alternatively, Manzini and Mariotti (2007) and Masatlioglu *et al.* (2012) consider models of deterministic consideration of prizes.

$$\lambda_l = \eta_1,$$
$$\lambda_r = (1 - \eta_1)\eta_2.$$

To illustrate failure of the NIAS conditions, note that substituting this data into the NIAS inequality for action L produces,

$$\eta_1 \geq \frac{(1 - \mu_l)\eta_2}{\mu_l + (1 - \mu_l)\eta_2},$$

which is clearly violated when η_1 and μ_l are small.

Note that in this simple example it is hardly surprising that the stochastic consideration set model produces failures of updating. After all, it is trivial for Bayesians to select the best prize if they examine just one prize: if the prize is x_1 it should be chosen, otherwise the unseen prize should be chosen.

4.2. *Stochastic Consideration of Actions*

Rubinstein and Salant (2006) describe a choice procedure that consists of searching a fixed number of positions in a list and then selecting the best searched option. In our language, such a procedure produces deterministic consideration of actions. To be more precise, let the action of choosing the first position in a list be a_1. Searching the first position in a list is equivalent to determining the prize associated with action a_1. Caplin and Dean (2011) study analogous models of search, but allow for stochasticity in the order of search. This gives rise to stochastic consideration of actions.

To study stochastic consideration of actions, we amend the earlier analysis by interpreting η_1 as the probability that action L is considered. The data produced by this theory reflect the fact that action L is selected if and only if it is considered and gives the best prize of the considered actions,

$$\lambda_l = \eta_1,$$
$$\lambda_r = \eta_1(1 - \eta_2).$$

Substituting this data into the NIAS inequality for action L produces a constraint on the consideration of the other action,

$$\eta_2 \geq \frac{1 - 2\mu_l}{1 - \mu_l},$$

which is clearly violated when η_2 and μ_l are small.

4.3. *Logit Demand*

One of the most important models of discrete choice is the logit model. This form of demand is sometimes motivated as resulting from imperfect cognition (McKelvey and Palfrey, 1995). However, we show the standard version of this model it is not consistent with BEU maximisation. To establish this, we simply analyse whether or not the associated stochastic choice data satisfy the NIAS inequalities.

Logit demand, which arises when errors follow an extreme value distribution, produces the following data,

$$\lambda_l = \frac{e^{\frac{1}{\kappa}}}{1+e^{\frac{1}{\kappa}}},$$
$$\lambda_r = \frac{1}{1+e^{\frac{1}{\kappa}}},$$

where $\kappa > 0$ is a parameter of the distribution. By way of interpretation, when the good prize is on the left, it is seen as being on the left with a probability that is increasing in how much better it is than the prize on the right. In this sense, rewards shrink stochastic errors.

To illustrate failure of the NIAS conditions, note that substituting this data into the NIAS inequality for selecting the good on the left produces

$$e^{\frac{1}{\kappa}} \geq \frac{(1-\mu_l)}{\mu_l},$$

which is violated whenever $\kappa > 1/\ln[(1-\mu_l)/\mu_l]$. To understand why logit demand cannot be rationalised in a Bayesian manner, note that prior beliefs play no role in determining stochastic choice. With a sufficiently uneven prior, logit demand is inconsistent with the NIAS inequalities.

On the other hand, Matějka and McKay (2011) show that a more generalised form of logit demand can be produced with rational inattention theory (Sims, 2003). In this example, that logit demand would be:

$$\lambda_l = \frac{e^{\frac{1}{\kappa}}}{e^{\frac{1}{\kappa}}-1} - \frac{e^{\frac{1}{\kappa}}}{\mu_l\left(e^{\frac{2}{\kappa}}-1\right)};$$
$$\lambda_r = \frac{(1-\mu_l)e^{\frac{1}{\kappa}}\left(\mu_l - \frac{e^{\frac{1}{\kappa}}-1}{e^{\frac{2}{\kappa}}-1}\right)}{\left(e^{\frac{2}{\kappa}}-1\right)}.$$

Substituting this data into the NIAS inequality for selecting the good on the left produces

$$\frac{e^{\frac{2}{\kappa}}-e^{\frac{1}{\kappa}}}{e^{\frac{2}{\kappa}}-1} \geq \frac{1}{2},$$

which is always satisfied because as κ approaches 0, the limit of the left-hand side is 1 and, as κ approaches ∞, the limit of the left-hand side is $1/2$. Thus, not surprisingly, this form of logit demand satisfies the standard assumptions.[7]

[7] Because of the symmetry in pay-offs and attentional costs, we need to consider only one of the constraints.

5. Unobservable Prior Probabilities or Subjective Priors

The characterisation above treats prior probabilities as observable and commonly understood between the outside observer and the decision-maker. However, our main result goes through even when the prior probabilities are not observable or we allow for the decision-maker to have subjective priors. If we retain also the assumption that all states are viewed as possible, so that $\mu \in \Gamma^I$, the definition of a BEU representation remains essentially unchanged and the characterisation result is precisely as before.[8]

While adding the prior to the set of unobservables adds an additional degree of freedom to match observables, the NIAS inequalities for this case remain restrictive. We illustrate the remaining restrictions in a relatively simple example.

EXAMPLE 1. *Consider an example with actions a, b, $c \in A$, prizes x, $y \in X$, and states ω, υ, $\tau \in \Omega$. Let prize x be given by action a in state ω, action b in state υ and action c in state τ, and prize y otherwise, so that $\omega(a) = \upsilon(b) = \tau(c) = x$ and $\omega(b) = \omega(c) = \upsilon(a) = \upsilon(c) = \tau(a) = \tau(b) = y$. Finally, consider the following data set:*

$$(q^a_\omega, q^b_\omega, q^c_\omega) = \left(\frac{1}{3}, \frac{2}{3}, 0\right);$$
$$(q^a_\upsilon, q^b_\upsilon, q^c_\upsilon) = \left(0, \frac{1}{3}, \frac{2}{3}\right);$$
$$(q^a_\tau, q^b_\tau, q^c_\tau) = \left(0, \frac{2}{3}, \frac{1}{3}\right).$$

For this data, there does not exist a $U : X \to \mathbb{R}$ and $\mu \in \Gamma^I$ that satisfy the NIAS inequalities. To confirm this, note first that the NIAS inequalities in this case require that $U(x) > U(y)$. This follows as with $q^a_\upsilon = q^a_\tau = 0$, the NIAS inequality for switching action a to action b asserts,

$$\frac{\mu_\omega}{3} U(x) \geq \frac{\mu_\omega}{3} U(y).$$

Given that $\mu_\omega > 0$, this requires that $U(x) \geq U(y)$. That the inequality must be strict follows from the fact that if it is not, then no NIAS inequalities hold strictly as the definition requires.

To complete our demonstration of non-existence, we consider the NIAS inequalities for switching from action b to action c, and from action c to action b. Noting that there is no loss of generality in setting $U(x) = 1$ and $U(y) = 0$, the required inequalities can be written as

$$\mu_\upsilon q^b_\upsilon \geq \mu_\tau q^b_\tau \quad \text{and} \quad \mu_\tau q^c_\tau \geq \mu_\upsilon q^c_\upsilon.$$

Substitution yields,

$$\frac{\mu_\upsilon}{3} \geq \frac{2\mu_\tau}{3} \quad \text{and} \quad \frac{\mu_\tau}{3} \geq \frac{2\mu_\upsilon}{3}.$$

Together, these inequalities require $\mu_\upsilon = \mu_\tau = 0$, contradicting $\mu \in \Gamma^I$.

[8] An open issue is how to characterise choices of a decision-maker with a subjective state space that may differ from that envisioned by the outside observer. It is important in such cases that the decision-maker's state space and prior allow for actions to yield the prizes they have been observed in practice to yield.

6. Strategic Analogues and Additional Applications

There are several strategic analogues to our model, two of which we discuss in this section.

6.1. *Bayes Correlated Equilibrium*

Bergemann and Morris (2013*a*) study games of incomplete information in which players may or may not have access to private information about the state. For the case of one player, they provide a definition of Bayes correlated equilibrium (BCE) for an arbitrary game G and 'experiment' S, where a game G is a triple (A, U, μ) and an experiment S is a set of signals T and an 'information structure' π. They define BCE with an 'obedience' condition on the player's decision rule σ: $T \times \Theta \rightarrow \Delta(A)$, where Θ is the set of pay-off relevant states.

The authors show that σ is a BCE of (G, S) if and only if, for some 'expansion' S^* of S, σ is a Bayes–Nash equilibrium (BNE) of (G, S^*), where expansion places an ordering on the informativeness of experiments. In other words, they show that if the player is playing a BCE for the game G with some information given by S, it is as if they are playing a BNE for game G with additional information beyond S given by S^*.

If there are no restrictions on the player's possible information, it is analogous to having the initial experiment S be completely uninformative, so that it contains only one signal. Such an S is called the 'null' experiment and is denoted by $\underline{S}$. For the null experiment, the decision rule reduces to a function σ: $\Theta \rightarrow \Delta(A)$, which is the same observable content as our model, so the NIAS conditions can be applied to this function.

Not surprisingly, the conditions for the existence of a BNE in a one player game when information is entirely unobservable (for the null experiment) are identical to the conditions for a BEU representation. It is immediate from their obedience condition that σ is a BCE of $(G, \underline{S})$ for some non-trivial U if and only if U satisfies the NIAS inequalities.

The chief difference between the approach that we take and that of Bergemann and Morris is around the observal of utility.[9] They treat utility functions as known and analyse possible equilibrium patterns of behaviour. We treat the data as given and infer utilities when these data satisfy conditions consistent with equilibrium play. Thus, our approach can be used to recover bounds on utility functions in the strategic setting just as it does in the decision theoretic setting. The distinction is that the resulting NIAS inequalities are joint restrictions on players' utility functions rather than restrictions that apply to each individual separately.

6.2. *Bayesian Persuasion*

Kamenica and Gentzkow (2011) determine necessary and sufficient conditions characterizing when a sender can benefit from sending a signal to a receiver who

[9] However, Bergemann and Morris (2013*b*) consider inference of utility functions beyond the linear quadratic normal framework.

takes a non-contractible action that impacts the utility of both parties. For the finite action case, the NIAS inequalities give the set of receiver actions that are possible under some signal choice and some receiver utility function. This statement reflects two differences between our setting and theirs. First, their action space is infinite rather than finite. Second, they treat the utility function as known.

6.3. *Additional Applications*

The centrality of BEU maximisation makes the NIAS inequalities of wide applicability. Caplin and Martin (2013*a*) apply the NIAS inequalities to show that observed data from a laboratory experiment are consistent with a general form of rational inattention. Caplin and Martin (2013*b*) extend the NIAS inequalities to categorise observed framing effects as having resulted from changes in perception or changes in utility. In a laboratory experiment, they find that explaining default effects requires distortions in utility. Caplin and Dean (2014) consider a more general environment, and find that in addition to the NIAS inequalities, a 'no improving attention cycles' (NIAC) condition characterises a general form of rational inattention theory. In a laboratory experiment involving a perception task, they find that subjects conform to both conditions in many, but not all, circumstances. Martin (2012) uses the NIAS inequalities to show that play in a strategic pricing experiment is consistent with rational inattention to quality.

New York University
Paris School of Economics

Submitted: 30 September 2013
Accepted: 8 January 2014

References

Afriat, S. (1967). 'The construction of a utility function from demand data', *International Economic Review*, vol. 8(1), pp. 67–77.

Bergemann, D. and Morris, S. (2013*a*). 'The comparison of information structures in games: Bayes correlated equilibrium and individual sufficiency', mimeo, available at http://dirkbergemann.commons.yale.edu/ (last accessed: 8 January 2014).

Bergemann, D. and Morris, S. (2013*b*). 'Robust predictions in games with incomplete information', *Econometrica*, vol. 81(4), pp. 1251–308.

Caplin, A. and Dean, M. (2011). 'Search, choice, and revealed preference', *Theoretical Economics*, vol. 6(1), pp. 19–48.

Caplin, A. and Dean, M. (2013). 'The behavioral implications of rational inattention with Shannon entropy', Working Paper 19318, NBER.

Caplin, A. and Dean, M. (2014). 'Revealed preference, rational inattention, and costly information acquisition', Working Paper 19876, NBER.

Caplin, A., Dean, M. and Martin, D. (2011). 'Search and satisficing', *American Economic Review*, vol. 101(7), pp. 2899–922.

Caplin, A. and Martin, D. (2013*a*). 'Defaults and attention: the drop out effect', Working Paper 17988, NBER.

Caplin, A. and Martin, D. (2013*b*). 'A revealed preference approach to attention and framing', mimeo available at, http://www.martinonline.org/daniel/ (last accessed: 8 January 2014).

Kamenica, E. and Gentzkow, M. (2011). 'Bayesian persuasion', *American Economic Review*, vol. 101(6), pp. 2590–615.

Manzini, P. and Mariotti, M. (2007). 'Sequentially rationalizable choice', *American Economic Review*, vol. 97(5), pp. 1824–40.

Manzini, P. and Mariotti, M. (2014). 'Stochastic choice and consideration sets', *Econometrica*, vol. 82(3), pp. 1153–76.
Martin, D. (2012). 'Strategic pricing with rational inattention to quality', mimeo, available at, http://www.martinonline.org/daniel/ (last accessed: 8 January 2014).
Masatlioglu, Y., Nakajima, D. and Ozbay, E. (2012). 'Revealed attention', *American Economic Review*, vol. 102 (5), pp. 2183–205.
Matějka, F. and McKay, A. (2011). 'Rational inattention to discrete choices: A new foundation for the multinomial logit model', mimeo, available at http://ideas.repec.org/p/cer/papers/wp442.html (last accessed: 8 January 2014).
McKelvey, R. and Palfrey, T. (1995). 'Quantal response equilibria for normal form games', *Games and Economic Behavior*, vol. 10(1), pp. 6–38.
Richter, M. (2011). 'Choice theory with equivalence', mimeo, available at, https://sites.google.com/site/richtereconomics/ (last accessed: 8 January 2014).
Rubinstein, A. and Salant, Y. (2006). 'A model of choice from lists', *Theoretical Economics*, vol. 1(1), pp. 3–17.
Rubinstein, A. and Salant, Y. (2011). 'Eliciting welfare preferences from behavioural data sets', *Review of Economic Studies*, vol. 79(1), pp. 375–87.
Salant, Y. and Rubinstein, A. (2008). '('A', 'f'): Choice with frames', *Review of Economic Studies*, vol. 75(4), pp. 1287–96.
Sims, C.A. (2003). 'Implications of rational inattention', *Journal of Monetary Economics*, vol. 50(3), pp. 665–90.
Tyson, C.J. (2008). 'Cognitive constraints, contraction consistency, and the satisficing criterion', *Journal of Economic Theory*, vol. 138(1), pp. 51–70.