ESRI

Working Paper No. 536

July 2016

The Surplus Identification Task and Limits to Multi-Attribute Consumer Choice

Peter D. Lunn[a,c], Marek Bohacek[a,b] and Féidhlim McGowan[a]*

Subsequently published as a Journal Article 'The surplus identification task and limits to multiattribute consumer choice' Journal of Experimental Psychology: Applied, Vol. 26(2), pp. 312–338 DOI: https://doi.org/10.1037/xap0000252

Abstract: We present a novel experimental method for investigating consumer choice. The Surplus Identification (S-ID) task is inspired by studies of detection in perceptual psychophysics. It employs a forced-choice procedure, in which participants must decide whether a novel product is worth more or less than the price at which it is being offered, that is, whether there is a positive or negative surplus. The SI-D task reveals how precision, bias and learning vary across attribute and price structures. We illustrate its use by testing for cognitive capacity constraints in multi-attribute choice in three separate experiments, with implications for models of bounded rationality and rational inattention. As the number of product attributes rises from one to four in the S-ID task (Experiment 1), participants cannot integrate additional information efficiently and they display systematic, persistent biases, despite incentivised opportunities to learn. Experiment 2 demonstrates how the S-ID task can be used to track learning and serves as a robustness check for the findings of Experiment 1. Experiment 3 adapts the S-ID task to test accuracy of surplus identification when multiple attributes are perfectly correlated. The S-ID task also has the potential to test multiple aspects of consumer choice models and to test specific hypotheses about the cognitive mechanisms behind surplus identification.

**Corresponding Author: pete.lunn@esri.ie*

Acknowledgements: These experiments were conducted as part of PRICE Lab, a research programme funded by four economic regulators in Ireland: the Central Bank of Ireland, Commission for Energy Regulation, Competition and Consumer Protection Commission and ComReg. The authors are especially grateful to Jason Somerville, Michael Morgan and Josh Solomon for assistance with the experimental environment and design, to Stephen Lea for illuminating discussions, and to audiences at the Economic and Social Research Institute, Network for Integrated Behavioural Sciences for feedback.

a The Economic & Social Research Institute, Dublin
b Institute for Neuroscience, Trinity College, Dublin
c Department of Economics, Trinity College, Dublin

The Surplus Identification Task and Limits to Multi-Attribute Consumer Choice *

Peter D. Lunn[⋆ ‡], Marek Bohacek[⋆ †], and Féidhlim McGowan[⋆]

[⋆]Economic and Social Research Institute

[†]Institute for Neuroscience, Trinity College Dublin

[‡]Department of Economics, Trinity College Dublin

July 8, 2016

*These experiments were conducted as part of PRICE Lab, a research programme funded by four economic regulators in Ireland: the Central Bank of Ireland, Commission for Energy Regulation, Competition and Consumer Protection Commission and ComReg. The authors are especially grateful to Jason Somerville, Michael Morgan and Josh Solomon for assistance with the experimental environment and design, to Stephen Lea for illuminating discussions, and to audiences at the Economic and Social Research Institute, Network for Integrated Behavioural Sciences for feedback.

Abstract

We present a novel experimental method for investigating consumer choice. The Surplus Identification (S-ID) task is inspired by studies of detection in perceptual psychophysics. It employs a forced-choice procedure, in which participants must decide whether a novel product is worth more or less than the price at which it is being offered, that is, whether there is a positive or negative surplus. The SI-D task reveals how precision, bias and learning vary across attribute and price structures. We illustrate its use by testing for cognitive capacity constraints in multi-attribute choice in three separate experiments, with implications for models of bounded rationality and rational inattention. As the number of product attributes rises from one to four in the S-ID task (Experiment 1), participants cannot integrate additional information efficiently and they display systematic, persistent biases, despite incentivised opportunities to learn. Experiment 2 demonstrates how the S-ID task can be used to track learning and serves as a robustness check for the findings of Experiment 1. Experiment 3 adapts the S-ID task to test accuracy of surplus identification when multiple attributes are perfectly correlated. The S-ID task also has the potential to test multiple aspects of consumer choice models and to test specific hypotheses about the cognitive mechanisms behind surplus identification.

1 Introduction

Economic agents need to identify surpluses in order to make gains from transactions. How accurately individuals accomplish this is difficult to gauge empirically because preferences and, consequently, consumer surpluses are generally unobservable. We introduce a novel empirical method, the Surplus Identification (S-ID) task, which brings the size of surpluses under experimental control and can be used to test and refine micro-economic models of consumer choice.

Previous empirical work has established numerous examples of consumers failing to choose optimally and therefore losing or missing out on surplus.[1] To generalise, these studies involve products or markets in which the good is essentially homogenous. This allows researchers to infer that where consumers do not choose low-cost options they are also failing to identify surpluses accurately. Together with the expanding volume of experimental work on decision-making biases, such findings have prompted revised models of consumer choice that incorporate limits to consumers' capacity for processing information. While models of bounded rationality date back to Simon (1955), more recent models of rational inattention (Sims, 2003; Woodford, 2014), salience (Bordalo et al., 2013), focussing (Kőszegi and Szeidl, 2013) and "relative thinking" (Bushong et al., 2015) define explicit constraints motivated by empirical findings. The first two are based on limits to the volume of information that can be be processed, the latter three on systematic re-weighting of product attributes across contexts.

This paper investigates cognitive constraints on surplus identification from an alterna-

[1]Relevant studies include: Wilson (2010) for residential electricity; Woodward and Hall (2012) for mortgages; Barber et al. (2005) and Choi et al. (2010) for mutual funds; Agarwal et al. (2009) and Agarwal et al. (2015) for credit products; Lambrecht and Skiera (2006) for broadband internet; Grubb (2009) for mobile telephony; and Bronnenberg et al. (2015) for branded medicines.

tive perspective, inspired by empirical techniques used to study perception. We consider surplus identification as a skill, like recognising or classifying objects. The latter requires immediate perceptions of an object's attributes to be compared with stored memories, generating an overall perception of what the object is. The stored information, which must be abstracted from previous perceptions and feedback, determines a mental mapping from the bundle of immediate perceptual cues to objects in memory. Mostly, people perceive objects correctly; sometimes they make mistakes. Surplus identification is a strikingly similar skill, albeit generally less perceptual and more cognitive. Immediate perceptions of a good's attributes, including its price, must be compared with stored memories. The stored information determines a mapping from the bundle of perceived attributes to an overall judgement of how much surplus is available. A standard model of consumer choice based on a utility function defined over attributes and prices is one way to model the information storage, mapping process and consequent surplus identification.

In empirical studies of perception, a central principle is to gain complete experimental control over the environment in which perception takes place. For the investigation of many perceptual capabilities, this process is relatively easy, because the objective state of, say, the visual, auditory or haptic environment can be perfectly manipulated by experimenters. Responses of experimental participants can be compared with the objective characteristics of the stimulus. Single dimensions of the perceptual environment can be altered and observers' sensitivity to change can be measured with precision. In such controlled environments, psychophysicists have used forced-choice procedures to make substantial progress in understanding how immediate percepts are mapped onto internal representations, revealing the limits of human perceptual ability (Macmillan and

Creelman, 2004). Probabilities of discrete choices, usually binary, are measured as a function of variables in the perceptual environment.

The S-ID task adapts this approach to investigate the ability to identify a surplus. Participants are introduced to a class of product with attribute magnitudes determined precisely by a computer. They are given a financial incentive to adopt a pre-determined preference function defined over attribute magnitudes and prices, which therefore governs the size of the monetary surplus for each instantiation of the product. The participant undertakes repeated forced-choice trials with feedback, with the size of the surplus ranging from large and negative to large and positive. On each trial the participant decides whether the surplus is positive or negative. The data reveal how the probability of detecting the surplus varies with its size, permitting precise estimation of the surplus required for reliable detection and of any bias toward overestimation or underestimation. The method also allows the time-course of learning to be tracked.

Thus, the S-ID task is incentive compatible, provides experimental control over exposure to the product and can measure the precision and bias of surplus identification as a function of any aspect of this exposure. As such, it offers opportunities for rigorous empirical investigation of specific cognitive constraints that bind consumers.

This paper introduces the method and demonstrates its potential by applying it to reveal how the detection of surpluses is affected by the number of attributes in the preference function. We show that surplus identification is surprisingly approximate and subject to systematic bias, with modest potential for learning. As the number of attributes rises from one to four, attribute information is not integrated with statistical efficiency and only large surpluses are detected reliably. These findings have implications for consumer choice models, which since Lancaster (1966) define preferences over multidimensional attribute spaces.

Section 2 explains how the S-ID task works, describing its key features and advantages. Section 3 applies the S-ID task to investigate the cognitive limitations of consumer choice as the number of product attributes rises, presenting results from three experiments. Section 4 concludes and discusses the potential for the S-ID task to yield further insights.

2 Key Features of the S-ID task

According to (Macmillan and Creelman, 2004), while detection theory was developed to study the sensitivity of the visual and auditory systems, the contemporary user "...more typically is interested in memory, cognition, or systems for medical or nonmedical diagnosis." (p. xiii). This exemplifies the success and scope of the associated empirical techniques in measuring the limits of performance in decisions of various kinds. This section outlines how we have adapted these techniques to develop the S-ID task. We are not aware of any previous attempt to apply them to consumers' abilities to detect surpluses.

2.1 Repeated forced-choice procedure

The S-ID task reduces surplus identification to a simple forced-choice task, repeated over multiple trials. On each trial, the participant sees a product and a price on a computer screen and must decide whether there is a positive or negative surplus, or, equivalently, whether the product is worth more or less than the price.[2] We refer to the price presented

[2]Although it is feasible to increase the number of alternatives, the experiments we report here presented only one product and an associated price, making it an adapted form of the classic two-alternative

on trial t as the "displayed" price, P_t^d, and the monetary value of the product as the "product price", $P_t = P_t^d + \Delta_t$, where the product possesses a surplus of Δ_t. Participants respond by pressing one of two buttons on a response box to indicate whether they judge Δ_t to be negative or positive.

The product price is an objective function of the product's attributes, $P_t = f(x_{1t}, ..., x_{kt})$, set by the experimenter. Participants are initially shown multiple examples of products and prices designed to assist them in learning the attribute-price relationship. After each response they receive feedback, consisting of an auditory beep for an incorrect response and the presentation of the product price, which remains beside the product until the participant presses a "next" button.

An important feature of the S-ID task is that participants are given an unambiguous incentive to respond as accurately as possible and, therefore, to adopt the preference function that determines the product price. In the experiments reported here, we used a tournament incentive such that one-in-ten participant's stood to win €50. This prize is substantial by the standards of pay-offs in experimental economics; it is a clear incentive to learn and to apply the pre-determined preference function. In this way, the SI-D task simulates the process by which a consumer encounters a new product and learns through experience and feedback to compare its worth against prices.

While simplified, the forced-choice procedure of the S-ID task is more akin to consumer choice than, for instance, pricing, valuation or willingness-to-pay tasks. In these other tasks, experimental participants or survey respondents are required to provide figures to match their judgements or preferences. By contrast, consumers generally make choices

forced-choice (2AFC) task used to study perception.

without the need to generate numbers. The same is true of the S-ID task, in which the choice is discrete and all participants must do is to decide whether what is before them represents a good or a bad deal.

2.2 "Hyperproducts"

Since the intention is to impose preferences via a clear incentive to adopt a pre-determined function, the S-ID task uses a novel product about which participants are unlikely to have strong pre-existing preferences or views regarding what determines value. To make the task engaging, the product is also designed to be intuitively valuable and pleasant to look at. Furthermore, because we are interested in how accurately agents integrate information when looking for surpluses, the attributes of the product consist of visual features that are known to be perceived with high accuracy.

To illustrate these properties of the SI-D task, Figure 1 provides a diagram of the product that formed the basis for the present experiments, which was a golden egg. This is an object that no experimental participant would be likely ever to have valued or traded, yet is intuitively valuable. The eggs were bright gold with anti-aliased edges and lines. They were generated in real time such that each had a unique pattern. The pleasantness of the images was ensured through extensive piloting. The product price could be determined by up to four attributes: the size; the fineness of surface texture; the circularity of an elliptical marking and the angle of a "hallmark" cross at the centre of the egg. These visual attributes were chosen based on studies of visual perception, which show that the relative magnitudes of size, texture (highest spatial frequency), elliptical

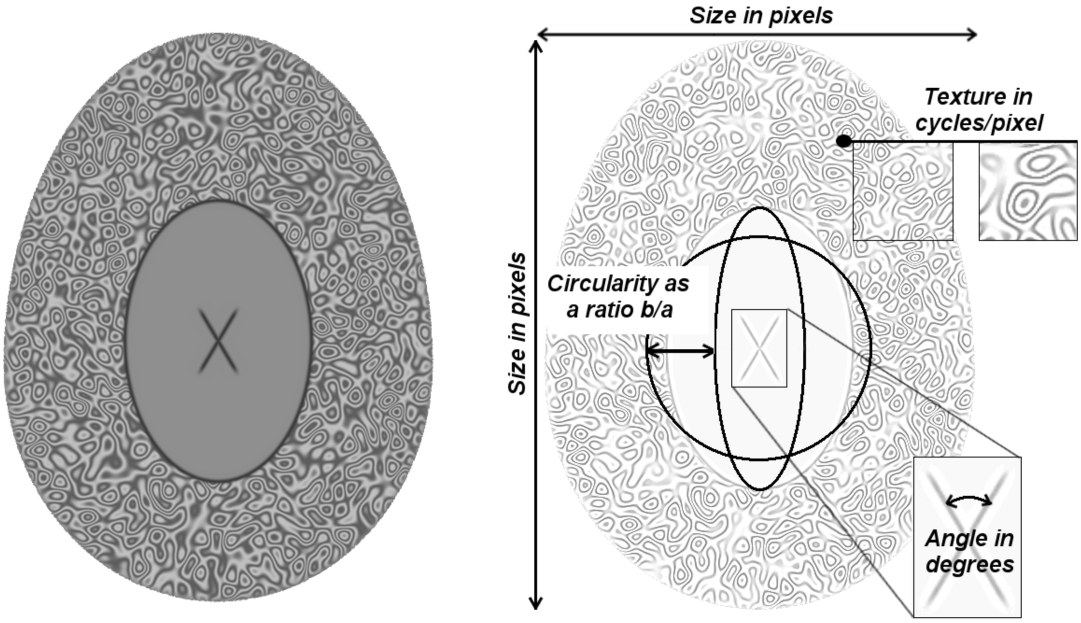

Figure 1: Golden egg "hyperproduct".

aspect ratio and angle can be discriminated with high accuracy .[3] Each attribute was pre-tested to ensure that this was the case when these visual stimuli were embedded within our eggs.

The use of visual attributes discriminated with high acuity is important. The aim is to test how accurately humans integrate information to identify a surplus, not the accuracy of human perceptual or pattern recognition systems. The attributes we employ keep imprecision from perception to a minimum. Furthermore, the aim, at least initially, is not to test integration of information via arithmetic or logical operations. Attribute information could, of course, be presented as abstract information, such as numeric values or categories. If so, over repeated trials participants might attempt to second guess the equation of the preference function, or perhaps test logical rules. These are interesting

[3]See Morgan (1991), Heeley and Buchanan-Smith (1990), Morgan (2005), Phillips and Todd (2010).

abilities for study that are likely to be involved in some consumer choices, especially where attributes are arithmetically or logically related (e.g., a percentage discount applied to a standard price, a minimum size to perform a function, etc.). The S-ID task can be adapted to investigate whether numeric and categoric attributes can be integrated more or less accurately than visual attributes. However, while there are clear exceptions, we contend that most consumer choice involves negotiating trade-offs without applying arithmetic or logical rules; consumers use their judgement to weigh product attributes and prices against each other. Employing visual attributes that are known to be discriminated with high acuity means that we can focus on the accuracy of such judgement. To learn and to apply the preference function participants must build up associations in memory, but attribute magnitudes must nevertheless be integrated afresh on each trial to assess the surplus.

We refer to our novel computerised products as "hyperproducts". The name reflects the fact that each product is uniquely defined in an attribute-price hyperspace. As such, its surplus is under complete experimental control.

2.3 Main empirical measures

Throughout this paper we employ two descriptive measures of the accuracy of surplus identification. The "just noticeable difference" (JND) and the "bias" are standard measures in detection theory. Both are generated by fitting the binary response data with a "psychometric function" that relates the size of the signal to be detected, in this case the surplus, to the probability of detection. The relationship is illustrated in Figure 2 using data from a single participant in Experiment 1. A logistic curve is fitted to the probability that the participant decides there is a positive surplus, as a function of the size of the surplus, measured as a proportion of the price range. The curve can be defined by two parameters: location and slope. For an unbiased participant, the logistic function will be centred at zero surplus. Significant deviation from zero indicates the bias. A positive bias equates to underestimation of surplus, because a positive surplus is need for the product and price to be judged equivalent; negative bias indicates overestimation. The slope of the logistic is a measure of the precision of performance, net of any bias. The JND is the size of surplus required to raise the probability of detection from 0.5 to 0.86, which equates to one standard deviation of a logistic distribution. Intuitively, it is the difference in surplus required for it to be detected with that level of reliability – hence "just noticeable".

Traditionally, researchers working on detection collect repeated measures of the JND and bias for a given task, then analyse changes in mean performance separately across individuals and experimental conditions. We adopt the more powerful econometric technique of fitting mixed effect logit (MEL) models for multiple individuals and conditions simultaneously, allowing for random variation in intercept and slope at the individual level. This method follows Moscatelli et al. (2012), who provide evidence for increased statistical power relative to the more traditional approach. Experimental conditions are included as independent variables in the model, permitting significance tests for their effect on the intercept and, when interacted with the surplus, slope of the psychometric function estimated for the sample as a whole. Further independent variables can be added in a similar way. For instance, a variable for the trial number can be specified to test whether the location and slope of the function are altered by learning as the ex-

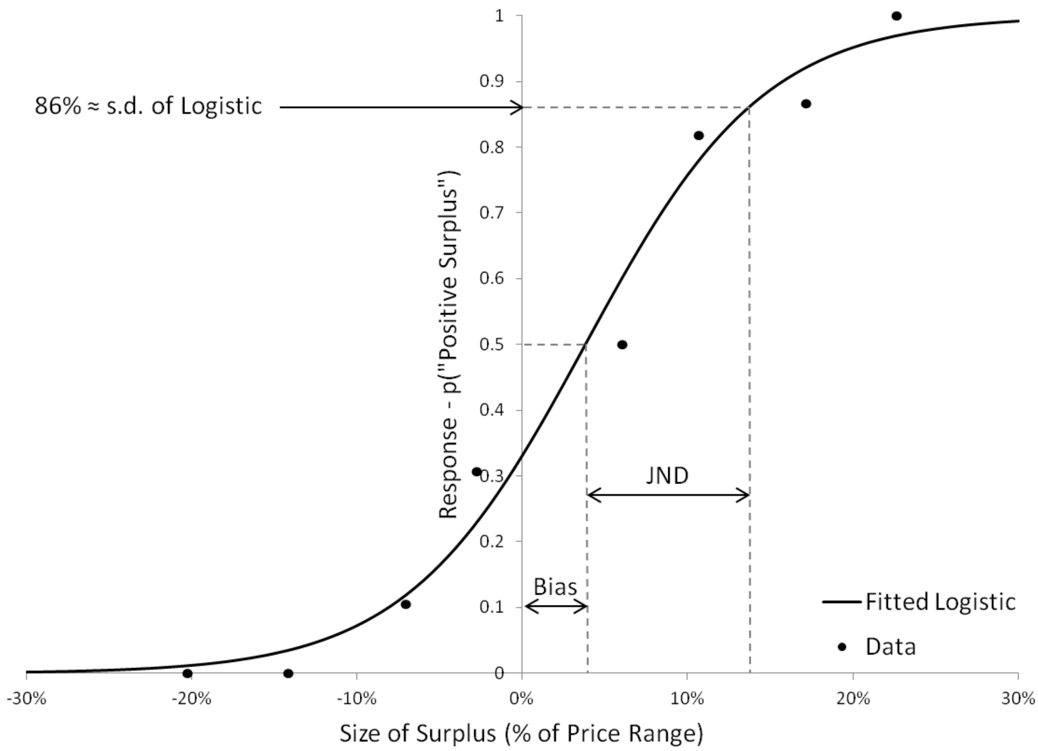

Figure 2: Psychometric function relating the size of the surplus to the probability of accurate detection. Main measures of performance are the JND, which reflects the precision of judgement, and the bias.

periment progresses. For ease of interpretation, average JNDs and biases for the sample can be computed from the model coefficients, according to the formulae provided in the Appendix.

The generation of these two separate measures encapsulates an important advantage of the S-ID task. Much empirical investigation of limitations in consumer choice involves measurement only of the extent to which contextual factors bias choice. The S-ID task can be used to measure behavioural biases, but has the additional capacity to reveal how those same factors affect the precision of consumer choice.

2.4 Potential objections

Before describing how we employed the S-ID task to investigate consumers' ability to integrate information from multiple attributes, it is worth considering some potential objections to using these techniques to study consumer choice.

Firstly, it could be contended that different cognitive constraints may affect choice when preferences are imposed on experimental participants through incentives, as in the S-ID task, than when preferences reflect purely subjective tastes. While we cannot conclusively rule this out, there are both conceptual and empirical reasons to doubt it. On a conceptual level, note that outcomes for consumers are often experienced as post-purchase feedback (e.g. how long a consumer durable lasts, the realised returns to an investment product, how easy an appliance is to use, etc), requiring them to integrate the feedback information into future assessments of surplus. While feedback in the S-ID task is more immediate, frequent and precise, there is no reason to suppose that a different cognitive mechanism processes it. Empirically, the rich data generated by the S-ID task allow us to test for biases documented in studies of purely subjective choices. Common biases imply common cognitive mechanisms.

Secondly, it might be argued that the S-ID task is really a perceptual judgement rather than a consumer choice task. The primary skill being tested is not perceptual, however. Rather, the S-ID task tests the ability to map attributes to prices and to trade them off against each other – a skill central to consumer choice. As explained above, visual attributes perceived with high acuity are employed for reasons of experimental control. Furthermore, while the three experiments described below used only visual attributes, in separate work we record little (if any) difference in performance across multiple experiments when visual attributes are replaced with numeric or categorical ones (Lunn and Somerville, 2016; Lunn et al., 2016).

Lastly, because the S-ID task entails multiple decisions over a relatively short period, it is possible that participants will not expend the cognitive effort that accompanies real consumer choices. However, participants are incentivised and permitted to respond in their own time. By recording response times, the experimenters can check whether participants take longer over more difficult trials – a characteristic of decision-makers aiming to optimise performance (Heitz, 2014). Moreover, we undertake a manipulation (in Experiment 2) that demonstrably increases effort with no effect on performance.

3 Application to Multi-Attribute Consumer Choice

We apply the S-ID task to investigate cognitive limitations associated with the number of product attributes. Since Lancaster (1966) standard micro-economic models of consumer choice have centred on the maximisation of utility in a multidimensional attribute-price space. In recent models that incorporate cognitive limitations, utility is assumed to be an additive function of weighted attribute magnitudes. Following the notation of Bordalo et al. (2013),

$$U(\mathbf{q}_k) = \sum_{i=1}^{m} \theta_i q_{ik} - \theta_p p_k \tag{1}$$

where each good k gives rise to a vector of m quality attributes $\mathbf{q}_k = (q_{1k}, q_{2k}, ..., q_{mk})$ and $(\theta_1, \theta_2, ..., \theta_m, \theta_p)$ are weights assigned to attributes and prices. Failures to maximise surplus arise where attributes and/or prices are inconsistently weighted across contexts. Motivated by empirical evidence that more "salient" attributes or prices are overweighted, Bordalo et al. (2013) propose that weights are a function of how the magnitude compares with the mean across the choice set. Magnitudes that "stick out" as large or small receive additional weight. Similarly, yet distinctly, Kőszegi and Szeidl (2013) propose that additional weight is given to attributes that vary more across the choice set, following empirical evidence for a "focussing illusion" by which agents overweight such attributes (Schkade and Kahneman, 1998). Bushong et al. (2015) present a contrasting model of how attribute ranges determine weightings.

In these models, cognitive constraints are implicit in the inability to weight attributes and/or prices consistently across contexts. Supporting empirical evidence suggests that there are at least some occasions when salience, focussing or range effects lead to biased choices and missed surplus. No limit is placed on the number of attributes in the preference function, m, and its relationship to the accuracy of surplus identification is unspecified. The S-ID task, however, permits a detailed empirical examination of this relationship. We measure how the accuracy of surplus identification is affected when the number of attributes in the preference function rises from one to four. This amounts to a test of fundamental cognitive constraints that afflict utility functions with the structure envisaged, rather than a test of the specific innovations designed to be consistent with biased attribute weightings.

Existing literature in experimental psychology is instructive but ambiguous with respect to the likely outcome. Since the seminal work of Miller (1956), it has been understood that absolute (as opposed to relative) judgements are subject to capacity limits. This evidence now includes: absolute identification experiments, in which participants match perceptual stimuli to ordered categories (Stewart et al., 2005); magnitude estimation studies, in which participants assign magnitudes to stimuli (Stevens, 1975); and judgement tasks variously described as "multiple-cue probability", "quantity" or "function" learning, in which participants guess the magnitude of an outcome variable from a set of cues (Lee and Yates, 1992).[4] Each of these experimental paradigms shares with the S-ID task the need to integrate information relating to the magnitudes of incommensurate scales. To generalise, results indicate significant imprecision and bias, even when the number of scales to be integrated is small.

Unlike the S-ID task, however, the above tasks lack incentive-compatible designs and, perhaps importantly, require participants to generate numeric responses to stimuli. When making choices consumers do not need to generate numbers for each surplus, but simply to make discrete choices over options. Reviewing multiple perceptual studies, Laming (1997) argues that the requirement to respond with numbers is likely to introduce sub-

[4] In a meta-analysis, Lee and Yates (1992) find that the correlation between the participant's guess and the outcome variable declines with the number of cues. The implications for the present study are unclear, however, since the relationships between cues and outcome variables are estimates of real world stochastic relationships that themselves vary in strength. Contrastingly, in the S-ID task, the attribute-price relationship is deterministic and under experimental control.

stantial error. Furthermore, Morgan et al. (2000) provide direct evidence that people can perform forced-choice discrimination of perceptual magnitudes against multiple internal standards, simultaneously, with minimal loss of precision compared to a relative judgement. Discrete choices may be more accurate than judgements that require people to generate numbers.

In this context, the studies described below demonstrate one of many potential applications of the S-ID task. Three experiments show how the accuracy of surplus identification relates to the number of product attributes, with implications for how cognitive constraints are built into consumer choice models. Moreover, because the task is incentive compatible and generates precise, quantitative estimates of how accurately consumers can assess surpluses, the experimental findings offer insights beyond current evidence from experimental economics and psychology.

3.1 Experiment 1

3.1.1 Method

Thirty-six participants were recruited by a market research company (15 male, 21 female; mean age=38, sd=12; 23 working, 13 not working). Each was paid €20 for participation and was told during recruitment and again on arrival that the most accurate performers, at least one in every ten, stood to win a €50 shopping voucher. After the experiment, prizes were mailed out to four participants with the highest mean ranking across conditions.

As described above, participants undertook an S-ID task involving a golden egg. On each trial they were presented with an egg and a displayed price. Their task was to decide whether there was a surplus, i.e., whether the monetary value of the egg was greater or less than the displayed price. This monetary value, the product price in Euro on trial t, was determined by an additive linear function:

$$P_t = 300 + \beta_1(A_{1t} - \bar{A}_1) + ... + \beta_m(A_{mt} - \bar{A}_m) \quad (2)$$

where $m \in \{1, 2, 3, 4\}$, A_{it} was the attribute magnitude of attribute i with mean magnitude $\bar{A}_i$ and β_i was the conversion factor for translating units of attribute i into Euro, with a mean product price of €300. Eggs ranged in value from €180 to €420. Attribute ranges were set to be approximately 26 JNDs, based on pilot studies in which participants discriminated each attribute individually when two eggs were presented side by side. That is, participants were able to discriminate approximately 26 different levels of each attribute reliably when seen simultaneously. The conversion factor, β_i, for each attribute was then set so as to match the attribute range to the price range.

Sessions proceeded as follows. Each participant initially undertook a learning phase and practice trials. They were first shown a series of eggs and corresponding prices based on a single attribute. They then completed 24 practice trials in this single-attribute condition. Next it was explained that they needed to consider a second attribute as well. They were shown examples with the first attribute held constant while the new attribute varied, then examples where both varied. They then completed 24 practice trials in the two-attribute condition. The three-attribute and four-attribute conditions were introduced in the same way. At the end of this learning and practice phase, participants had viewed 36 example eggs and completed 96 practice trials with feedback. There was then a break before the test phase, which consisted of four runs of 80 trials each. Participants proceeded at their own speed, with breaks between runs. The session, including breaks, lasted around one hour.

The order of test runs was counterbalanced across participants, such that one quarter completed the single-attribute condition first, one quarter the two-attribute condition and so on. There were 15 possible attribute combinations: four one-, six two-, four three-, and one four-attribute combination. These subconditions were also counterbalanced across participants.

The displayed price, P_t^d, product price, $P_t = P_t^d + \Delta_t$ on each trial were selected as follows. Attribute magnitudes were generated randomly from uniform distributions. The corresponding price was calculated. This determined the displayed price. The surplus, Δ_t, was added to obtain the product price. The relevant attributes were then increased or decreased to match this price, with proportions of the increase or decrease assigned across attributes at random, subject to attribute magnitudes and prices remaining within ranges. This included ensuring that for any given displayed price and product price, the opposite sign surplus, $-\Delta_t$, would also keep the product price within range. Hence, whatever the displayed price, the probability of a positive or negative surplus was always 0.5. If a price or attribute magnitude fell outside the specified range, the program began the process again. With this method, means and standard deviations of prices and attributes shown were approximately constant across conditions.

The surplus for each trial was chosen to be either positive or negative, Δ_t, with its precise magnitude controlled by two "1-up 2-down" staircases (Garcıa-Pérez, 1998), which adapted to previous performance. Following two correct identifications of surpluses of a given sign, the absolute magnitude of the next surplus of that sign was reduced; following one misidentifcation it was increased. Thus, after two correct responses the task became harder, while after one incorrect response it became easier. This adaptive procedure was employed because it tracks performance efficiently if participants improve during a run. The staircase converges to a level of difficulty where errors occur on around one-in-five trials, assisting estimation of performance.[5]

3.1.2 Results

Initially, separate psychometric functions were estimated for each of the 144 experimental runs (36 participants x 4 conditions). Six (4%) were discarded on the grounds that the participant failed to produce a monotonic function. These runs were excluded to improve point estimates, though our results are altered little by their inclusion. Estimated intercepts and slopes (coefficients on the surplus) were approximately normally distributed, with a modest correlation between the two. This supported the use of an MEL model with random effects on both coefficients, allowing for a correlation between the two.[6]

Table 1 presents a series of MEL models estimated on more than 11,000 binary responses. Model 1 reveals a number of effects. The positive and significant constant indicates a bias towards judging the surplus to be positive. Thus, participants tended to overvalue the product, with some variation according to the number of attributes. Unsurprisingly, the coefficient on the surplus is highly significant – the larger the absolute (positive or negative) surplus, the more likely participants were to identify it. The large and highly significant negative interaction of the surplus with the number of at-

[5] Note that participants were informed that the difficulty of the task adapted to performance and understood that they could not "game" the system by getting some trials wrong to obtain easier subsequent trials.

[6] Models were also re-estimated with fixed effects, i.e., separate dummy variables for each participant plus interactions with surplus, with minimal impact on estimated coefficients.

tributes shows that surplus identification became less precise as the number of attributes increased. These coefficients were used to calculate the JNDs presented in the left panel of Figure 3. Three aspects are noteworthy. First, even when the product had a single attribute, surplus identification was quite imprecise: a surplus equivalent to one-fifth of the price (or, equivalently, attribute) range was required to discriminate it reliably. Recall that the attribute range consisted of approximately 26 discriminable levels. Thus, once participants had to map these on to numeric prices to determine surpluses, judgement was far less precise. Second, precision declined dramatically as the number of attributes increased, with the JND climbing to over half the price-range once four attributes were in play. Third, Figure 3 compares performance with two-, three- and four- attribute products to that of a hypothetical "ideal" integrator of information. As shown in the Appendix, this curve would be generated by an observer who, given their precision when mapping a single attribute to a price, could integrate additional attributes with statistical efficiency, i.e. the variance of their estimates would be additive. The disparity between this curve and participants' performance is a clear impact of cognitive capacity constraints. Each time the amount of information increases, there is an additional loss of efficiency in the ability to identify the surplus.

Separate models were estimated in which the four category 'Attributes' variable was separated into 15 categories, corresponding to four different single-attribute conditions, six combinations of two attributes, four possible combinations of three attributes and the one four-attribute condition. The counterbalancing ensured that each combination was undertaken by the same number of participants. All two-attribute conditions resulted in higher JNDs than all single-attribute conditions and lower JNDs than all three-attribute combinations, which in turn all produced lower JNDs than the four-attribute condition.

Table 1: Mixed Effects Logit for p("positive"), Experiment 1

	(1)	(2)	(3)	(4)
Constant	0.294***	0.303***	0.336***	0.370***
	(0.077)	(0.080)	(0.079)	(0.085)
Attributes (Ref=One)				
Two	-0.027	-0.028	-0.047***	-0.080
	(0.068)	(0.068)	(0.069)	(0.075)
Three	-0.219***	-0.219***	-0.255***	-0.293***
	(0.066)	(0.066)	(0.067)	(0.072)
Four	-0.034	-0.035	-0.061***	-0.110
	(0.066)	(0.066)	(0.067)	(0.072)
Surplus	8.865***	8.861***	9.099***	9.889***
	(0.351)	(0.362)	(0.357)	(0.382)
*Attributes*Surplus*				
Two *Surplus	-3.793***	-3.794***	-3.833***	-4.735***
	(0.347)	(0.350)	(0.350)	(0.377)
Three*Surplus	-5.096***	-5.097***	-5.250***	-6.072***
	(0.332)	(0.332)	(0.336)	(0.362)
Four*Surplus	-6.151***	-6.151***	-6.312***	-7.172***
	(0.330)	(0.330)	(0.334)	(0.360)
Learning				
Trials 41-80		-0.017		
		(0.045)		
Trials 41-80		0.009		
		(0.180)		
Displayed Price				
Price			0.284***	0.641***
			(0.023)	(0.043)
Price*Surplus			0.090	-0.827***
			(0.115)	(0.309)
*Attributes*Price*				
Two*Price				-0.476***
				(0.064)
Three*Price				-0.514***
				(0.064)
Four*Price				-0.634***
				(0.066)
*Attributes*Price*Surplus*				
Two*Price*Surplus				1.151***
				(0.390)
Three*Price*Surplus				0.710*
				(0.372)
Four*Price*Surplus				0.993***
				(0.372)
Random effects parameters				
Var(Constant)	0.122	0.122	0.130	0.137
	(0.034)	(0.034)	(0.036)	(0.038)
Var(Surplus)	1.325	1.324	1.364	1.400
	(0.391)	(0.391)	(0.401)	(0.410)
Cov(Constant, Surplus)	0.105	0.105	0.122	0.121
	(0.084)	(0.084)	(0.088)	(0.091)
Observations	11,040	11,040	11,040	11,040
Individuals	36	36	36	36

Standard errors in parentheses. *** p<0.01; ** p<0.05; * p<0.1

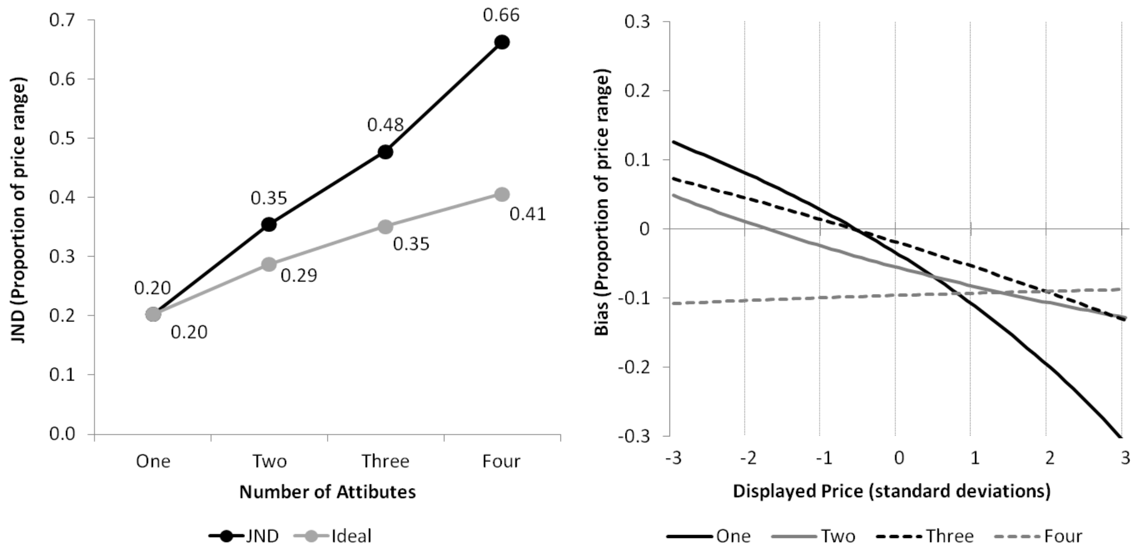

Figure 3: JND (left) and bias (right) in Experiment 1.

In short, imprecision was primarily determined by the number of attributes to be factored in.

Model 2 in Table 1 tests for learning by adding a dummy variable for whether the trial was in the first or second half of the experimental run. No significant learning effect was observed across the test trials.[7] Model 3 includes a standardised variable for the displayed price, to test whether performance varied over the price range. This reveals an unanticipated bias: the tendency to overestimate the value of the product was a function of whereabouts in the price range the comparison was made. Surplus tended to be underestimated at the lower end of the price range and overestimated at the higher end. Further investigation yielded Model 4, which shows that this bias also varied significantly with the number of attributes. Since these coefficients are not straightforward to interpret, the right panel of Figure 3 plots the implied bias across the price range. The general overestimation of surplus (negative bias) was small compared to the variation in bias by

[7] This remained true whether the variable for trial number was categorical, continuous, linear or non-linear. Any measurable learning was confined to the practice trials.

price and number of attributes. When a single attribute determined the surplus there was a strong tendency to underestimate surplus towards the bottom of the price range and to overestimate it towards the top. This effect moderated as the number of attributes increased, producing the anti-clockwise rotation of the curves apparent in Figure 3. We return to potential causes of this bias in the discussion of Experiment 2. [8]

The rich data generated by the S-ID task allow further explanatory variables to be added to the models to test for behavioural biases commonly recorded in subjective choice experiments. As argued above, the presence of such biases supports the argument that common cognitive mechanisms are involved in the S-ID task and consumer choice. Table 2 presents three such tests, using Model 4 of Table 1 as the base model. First, in choice experiments consumers find it difficult to ignore irrelevant attributes, a finding referred to as the "dilution effect" (Meyvis and Janiszewski, 2002). Model 5 introduces a variable for the product price being signalled by the subset of attributes (one, two or three) that participants needed to ignore during the relevant experimental run, but to factor in during some other runs. The significant coefficient reveals that participants were biased by irrelevant attributes – a dilution effect. Second, in choice experiments consumers are less likely to opt for products with extreme attribute magnitudes (Simonson and Tversky, 1992) or, similarly, more likely to choose products in which attributes are "balanced" (Chernev, 2005). For instance, consumers tend to choose options with two attributes of middling magnitudes over options with one good and one bad. Such findings are in keeping with the notion of diminishing returns to attributes and consistent with many preference functions commonly used in economic applications. Model 6 introduces a vari-

[8] A model with all 15 attribute combinations revealed that this gradient in the bias across the price range was more pronounced for all four single-attribute conditions than for any of the other 11 attribute combinations.

able that corresponds to the Gini coefficient calculated from the standardised attribute magnitudes; the higher the Gini coefficient, the more unbalanced the attributes. Model 6 shows this variable to be highly significant. Larger trade-offs between attributes lowered the probability that participants judged the surplus to be positive. Interestingly, larger trade-offs also reduced precision – a finding discussed further below. Third, choice experiments show that consumers facing a multi-attribute choice tend to place greater weight on more familiar attributes Hauser (2011) or attributes that they encounter first (Carlson et al., 2006). Model 7 tests whether participants placed greater weight on the first attribute they were introduced to, which was also the attribute they had to assess in the single-attribute condition (and was counterbalanced across participants). when this initial attribute had the highest magnitude of the of two, three or four, participants were more inclined to decide that the surplus was positive. Model 8 confirms that when it had the lowest magnitude of three or four, the bias reversed. This simple bias in favour of an initial attribute is consistent with choice experiments. Interactions with the number of attributes were non-significant and there was no effect on precision.

3.1.3 Discussion

Experiment 1 shows how the S-ID task can generate empirical measures of how accurately consumers can identify surpluses and how this varies when properties of the product change. The results imply cognitive limits to the ability to integrate attribute and price information. A single, plainly visible attribute could be compared only imprecisely against a numeric price. Subsequent related work (Lunn and Somerville, 2016) confirms, first, that across a number of different products, attributes and price ranges surpluses

Table 2: Additional tests, Experiment 1 (Variables added to Model 4, Table 1)

	(5)	(6)	(7)	(8)
Irrelevant Attributes				
Irrelevant Signal	0.095***			
	(0.026)			
Irrelevant Signal*Surplus	-0.031			
	(0.100)			
Attribute Balance				
Gini		-0.559***		
		(0.184)		
Gini*Surplus		-2.522***		
		(0.720)		
First Attribute				
Highest			0.349***	0.272***
			(0.055)	(0.075)
Highest*Surplus			-0.058	-0.268
			(0.206)	(0.271)
Lowest				-0.017**
				(0.073)
Lowest*Surplus				-0.384
				(0.267)
Observations	8,320	8,240	8,240	5,520
Individuals	36	36	36	36

Standard errors in parentheses. *** p<0.01; ** p<0.05; * p<0.1
Model 5, four-attribute condition excluded; Models 6 and 7, single-attribute condition excluded; Model 8, single- and two-attribute condition excluded

equivalent to one-fifth of the price (or attribute) range are required for individuals to identify them reliably; second, that this result generalises to nonlinear attribute-price relationships; and, third, that the downward sloping bias across the price range also generalises to other products, attributes and prices.

The main focus of the present experiments, however, is the impact of increasing the number of attributes beyond one. Based on the level of precision when comparing a single attribute to a price, additional attributes could not be processed with statistical efficiency. Imprecision in surplus identification rose sharply. With four attributes to take into account, participants needed surpluses of more than half the price range; effectively, they could just about tell a good product from a bad one. Judgements of surplus were also systematically biased. Three biases previously recorded in choice experiments were present in the S-ID task data, suggesting that common psychological mechanisms are involved in both. In addition, surplus was systematically underestimated and overestimated at the bottom and top of the price range, although this bias moderated as the number of attributes increased. There was no evidence of learning following initial familiarisation and practice trials.

The implied constraints are not perceptual but cognitive. Participants had no difficulty distinguishing multiple levels of each attribute, only mapping those percepts on to a numeric price range to assess surplus. The suggestion is that there are cognitive limitations in the ability to map incommensurate scales to each other, which is an essential skill for spotting surpluses.

3.2 Experiment 2

Experiment 2 was developed as a robustness check for the findings of Experiment 1 and demonstrates how the S-ID task can be used to track learning. Experiment 1 might have underestimated performance for several reasons. First, although participants indicated that they fully understood the task, it entailed a degree of complexity and not all participants may have fully comprehended the task. Second, the within-subject design required participants to engage with four different conditions. It is possible that this generated some interference between conditions. Third, although carefully piloted, the one-hour session (including break) may have resulted in fatigue or a diminution of effort, despite the financial incentive. Fourth, since hyperproducts are new to participants and data were collected in a single session, any capacity to learn over periods of days was negated.

Experiment 2 sought to discount these possibilities. It consisted of a surplus identification tournament held among the research staff at a national research institute. All participants were highly numerate, educated to postgraduate level in a quantitative social science, mostly economics, and conducted data analysis as part of their daily work. The risk that any of them misunderstood the task was negligible. Each repeated just a single condition, with either two, three or four attributes. The length of the sessions was halved to approximately 30 minutes, with three sessions per participant separated by at least a week. Finally, an unanticipated additional incentive was introduced to induce additional effort on the final experimental run.

3.2.1 Method

Methods were as for Experiment 1 with the following adaptations. Participants were 24 members (13 male, 11 female) of the research staff at a national research institute. They were pseudo-randomly assigned to a two-, three- or four-attribute condition. Those in the two- and three-attribute conditions were further assigned a subcondition with one of two attribute combinations (size-circularity or texture-angle; size-texture-circularity or size-texture-angle). Participants were not paid, but were told that the best performer would win a 50 voucher (which we judged by comparing precision against standardised distributions by condition from Experiment 1).

Following the learning phase, participants undertook three experimental runs of 64 trials. This pattern was repeated in each session, except that in the third session we introduced an unanticipated incentive manipulation. Once the first run was complete, participants were expecting two more runs but were instead told that, in fact, they needed to undertake just one more and, furthermore, that the participant who improved most relative to their previous performance would win a 50 voucher.

3.2.2 Results

The left panel of Figure 4 presents JNDs by number of attributes and session, calculated as for Experiment 1. There were no significant differences in performance between subconditions with different attribute combinations. In the first session, JNDs were slightly lower than for the sample of consumers in Experiment 1, somewhat more so for the participants assigned to the three-attribute condition. Overall, the JND of the median participant in Experiment 2 would have placed them at the 69^{th} percentile in Experi-

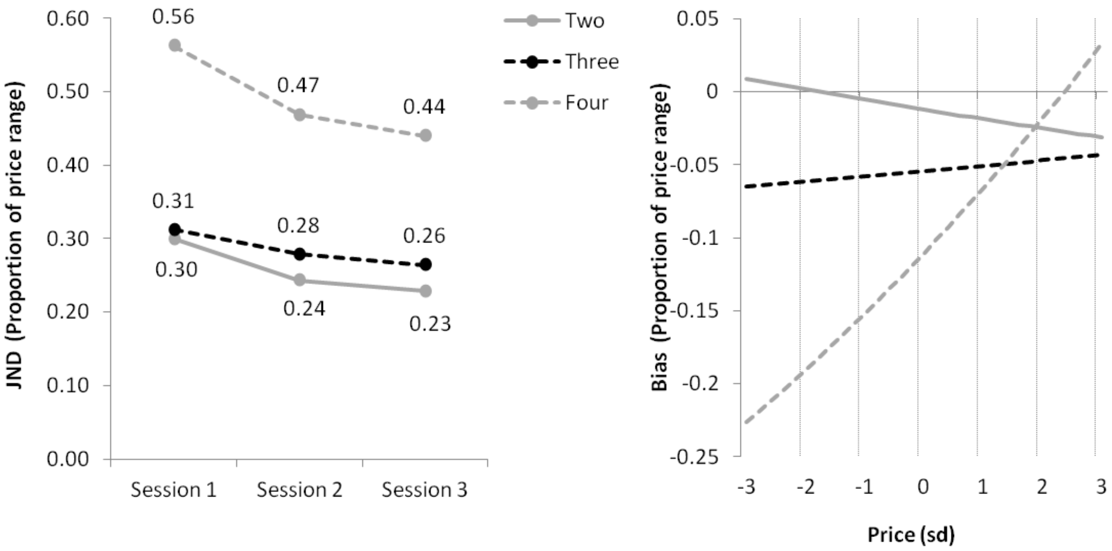

Figure 4: JND by session (left) and bias (right) in Experiment 2.

ment 1. There were small improvements in precision across sessions, following standard learning curves. In each case, the improvement between Sessions 1 and 2 was statistically significant, while the difference between Session 2 and 3 was not. Overall, following hundreds of trials with feedback spread over different days, this highly educated and numerate group still needed a surplus equivalent to a large proportion of the price range in order to identify it reliably.

The pattern of biases was more consistent and is shown for data pooled across sessions in Figure 4. Although the pattern is somewhat different from that recorded in Experiment 1, there are three commonalities: an overall tendency to overestimate surplus (negative bias); substantial variation across the price range; and an anticlockwise rotation in the relationship between price and bias as the number of attributes increased.

The effort manipulation was effective: response times for the final run of Session 3 were half a standard deviation longer than response times in the preceding run. This additional effort had no significant impact on performance, however. Half of the participants recorded a higher JND in the final run than in the penultimate one, the other half a lower JND. Biases were unaffected.

Table 1 presents three MEL models, estimated separately for the participants who undertook the two-, three- and four-attribute conditions.[9] The models confirm the significance of the pattern of improvement in precision and the bias across the price range just described. They also include a variable for the Gini coefficient, as in Experiment 1. The two coefficients are again highly statistically significant. Larger trade-offs between the attributes reduced the likelihood of deciding that the surplus was positive and reduced precision. Both effects were stable – interactions with session were non-significant .

The impact of the size of the trade-off on precision was large. Figure 5 plots estimated JNDs by Gini, calculated from the models in Table 3 and the equivalent models for Experiment 1. The differences in the range of the Gini reflect differences in the maximum possible Gini calculated for two-, three- and four-attributes. Despite the differences between the two experiments already described, Figure 5 confirms the two greatest influences on the precision with which surpluses could be reliably detected: the number of attributes that must be integrated into the decision and the size of the trade-off between their magnitudes.

3.2.3 Discussion

The S-ID task generates precise empirical measures of accuracy and learning, as well as rich data that offer insights into cognitive constraints operating in consumer choice. The highly numerate sample of individuals in Experiment 2 could identify surpluses somewhat

[9]Almost identical results are produced by a single model estimated for all 24 participants. Separate models are presented for ease of interpretation, since the single model contains multiple statistically significant three-way interactions.

Table 3: Mixed Effects Logit for p("positive"), Experiment 2

Number of Attributes	Two	Three	Four
Constant	0.694***	1.211***	1.890***
	(0.183)	(0.169)	(0.221)
Surplus	8.612***	8.375***	5.842***
	(0.647)	(0.937)	(0.792)
Session (Ref=Third)			
First	-0.221**	0.138	-0.018
	(0.104)	(0.104)	(0.095)
Second	0.183*	-0.135	0.150
	(0.106)	(0.103)	(0.097)
*Session*Surplus*			
First*Surplus	-1.995***	-1.064**	-0.823**
	(0.515)	(0.466)	(0.353)
Second*Surplus	-0.412	-0.419	-0.197
	(0.547)	(0.474)	(0.367)
Displayed Price			
Price	-0.047	-0.134***	-0.230***
	(0.039)	(0.039)	(0.041)
Price*Surplus	-0.225	-0.317	-0.218
	(0.235)	(0.217)	(0.184)
Attribute Balance			
Gini	-1.486***	-3.606***	-3.604***
	(0.179)	(0.468)	(0.450)
Gini*Surplus	-2.540***	-7.283***	-4.628***
	(0.858)	(1.982)	(1.651)
Random effects parameters			
Var(Constant)	0.172	0.065	0.052
	(0.094)	(0.039)	(0.032)
Var(Surplus)	0.888	3.792	0.364
	(0.607)	(2.044)	(0.264)
Cov(Constant, Surplus)	-0.091	-0.082	-0.030
	(0.166)	(0.202)	(0.068)
Observations	4,096	4,096	4,096
Individuals	8	8	8

Standard errors in parentheses. *** $p<0.01$; ** $p<0.05$; * $p<0.1$

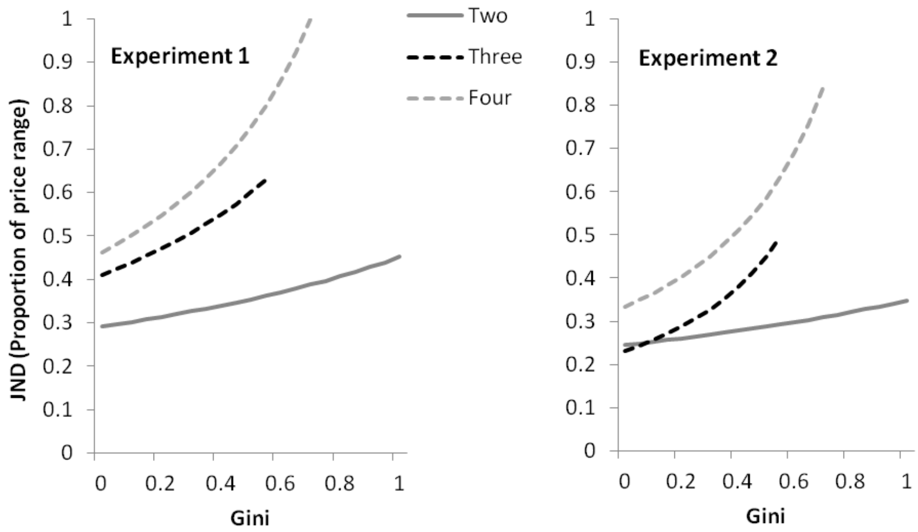

Figure 5: JND by Gini coefficient and number of attributes in Experiments 1 and 2.

more precisely than the sample of consumers in Experiment 1, and they displayed modest improvements over sessions consisting of hundreds of trials separated by days. Yet the central implications of Experiment 1 were reinforced by Experiment 2. The precision with which surpluses could be identified was limited by cognitive constraints relating to the number of attributes that had to be simultaneously processed and the size of the trade-offs between them. The psychological mechanisms involved also generated systematic biases across the price range that depended on the number of attributes in play.

3.3 Experiment 3

Experiment 3 demonstrates how the S-ID task can be used to test specific hypotheses about the mechanisms behind surplus identification. In Experiments 1 and 2, attribute magnitudes had a modest negative correlation, which was necessary to keep the product price within a defined range while varying the number of attributes. A statistical analysis of real products, however, reveals that attributes can be negatively correlated, positively correlated or, frequently, a mixture (Curry and Faulds, 1986). For instance, specifications of electronic goods, domestic appliances and cars often list multiple attributes intended to signal the same underlying level of quality, i.e. with strong positive correlation. Experiments 1 and 2 may underestimate capabilities if consumers can integrate information from multiple positively correlated attributes more efficiently than multiple negatively correlated attributes. Experiment 3 adapts the method to test the accuracy of surplus identification when multiple attributes are perfectly correlated. If consumers can integrate such information accurately, performance should improve substantially as attributes are added, because each provides additional information about the same underlying value.

Experiment 3 also permits tests of contrasting hypotheses in relation to the cause of the biases recorded in Experiments 1 and 2. One possible explanation for the underestimation of surplus at the bottom of the price range and overestimation at the top derives from the limited range of neural coding. Barlow (1961) posited that neural systems adapt to a range of inputs so as to disperse neural responses and, thereby, increase sensitivity to differences within the range. However, while such adaptation increases the ability to discriminate reliably between similar inputs, it distorts the mapping between the inputs and the neural code, biasing responses away from the centre of the range. The implication is a precision-bias trade-off: the system can increase the ability to discriminate only at the expense of bias away from the centre of the range (Summerfield and Tsetsos, 2015). This account is consistent with the bias for single-attribute products, but requires a second factor to explain the anti-clockwise rotation of the bias with increasing attributes. One possibility is that perceived attribute magnitudes are averaged, following some process of normalisation. Choice experiments show that consumers sometimes decide by averaging attributes, leading to the counterintuitive finding that an option that is good on one attribute and moderate on another can be judged inferior to one that is known only to be good on the first (Troutman and Shanteau, 1976; Weaver et al., 2012). With the additive preference function employed here, attribute averaging would result in the recorded rotation in the bias. This rotation could also be caused by uncertainty weighting, however. If, as the number of attributes increases, individuals become more uncertain of the product price implied by attribute information, judgements may be biased towards the mean. These two accounts offer different predictions when attributes are perfectly correlated. Attribute averaging predicts that the rotation in the bias should disappear, as all attributes signal the same product price. Uncertainty weighting predicts that the rotation should reverse, provided uncertainty is at least to some extent reduced by multiple signals of the same product price.

3.3.1 Method

There were 24 participants (10 male, 14 female). Each was paid a €25 participation fee and was told that at least one in every ten stood to win a €50 shopping voucher. Three vouchers were mailed out after the experiment was completed to the participants with the highest mean ranking across conditions.

Methods were as for Experiment 1, with the following modifications. While there were again four conditions defined by the number of attributes that signalled the product price, the preference function was no longer additive:

$$P_t = 300 + \beta_1(A_{1t} - \bar{A}_1) = ... = 300 + \beta_i(A_{it} - \bar{A}_i), i \in \{1, 2, 3, 4\} \tag{3}$$

Thus, the participant was presented with either one, two, three or four perfectly correlated attributes that all signalled the same product price. Where $i<4$, the attributes not signalling the price were held constant at the midpoints of their ranges. To help participants differentiate between conditions four colours of precious eggs were used (gold, silver, bronze and emerald). The relationship between the colour and the number of attributes was counterbalanced across participants. Since the task was easier to grasp than in Experiment 1, we employed 12 practice trials per condition, followed by four pseudo-randomised experimental runs of 72 trials.

A staircase procedure was employed in Experiments 1 and 2 because it adapts rapidly to performance and we anticipated the potential for large learning effects. Given the absence of such effects, we opted for an alternative standard procedure also employed in the study of perception, the "adaptive method of constant stimuli" (ACMS). ACMS adapts to performance more slowly but tends to be more enjoyable for participants because it mixes easier and harder trials. Unbeknownst to the participant, each run of 72 trials consisted of six blocks of twelve. Within a block, Δ_t corresponded to three positive and three equal and opposite negative surpluses with a constant separation, $\{5\delta$, 3δ, δ, -δ, -3δ, -$5\delta\}$, where δ was a proportion of the price range, presented in a random order. If a participant responded correctly on more than ten trials in a block, δ was reduced for the next block. If they responded correctly on eight or fewer, δ was increased. We did not anticipate that use of the ACMS procedure would have any material impact on the ability to identify surpluses.

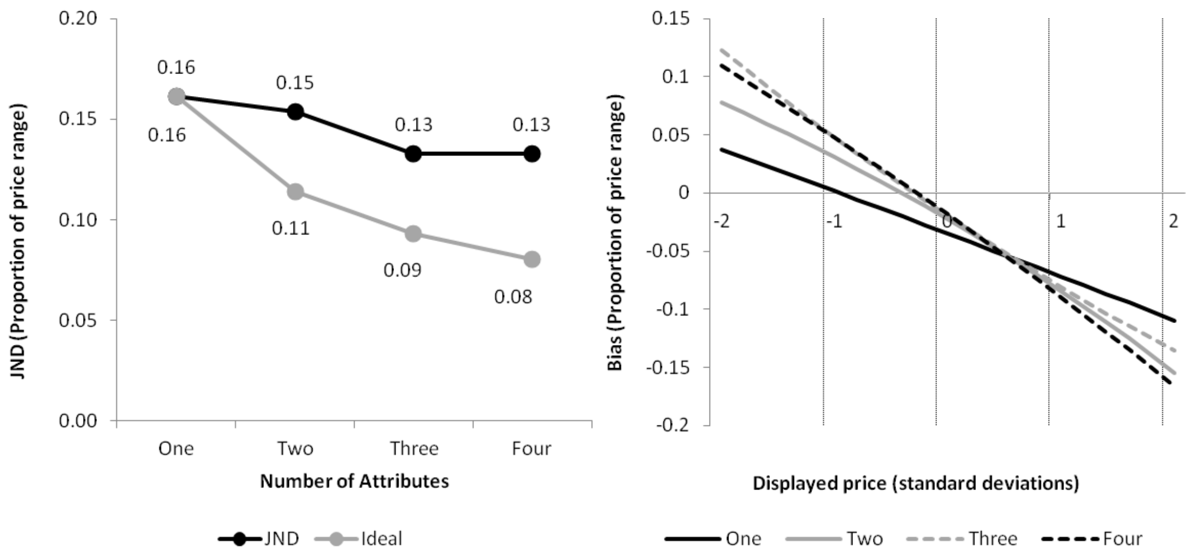

Figure 6: JND (left) and bias (right) in Experiment 3.

3.3.2 Results

Figure 6 presents JNDs and biases by number of attributes, calculated from an MEL model with the same specification as Model 4 in Table 2. The JND of 0.16 of the price range for a single-attribute hyperproduct is slightly better than the 0.20 recorded in Experiment 1. As in Experiment 1, JNDs for the two-, three- and four-attribute products are compared to those of a hypothetical "ideal" integrator of information. Imprecision was significantly reduced in the three- and four-attribute conditions relative to the single- and two-attribute conditions ($p<0.01$ and $p<0.05$ respectively), but there was no improvement between the three- and four-attribute conditions, while the overall improvement fell far short of ideal integration.

However, this increased precision came at the cost of greater bias. Consistent with a precision-bias trade-off and contrary to the attribute averaging account, the bias across the price range strengthened as the number of attributes (and precision) rose. The interaction between the standardised price and the number of attributes was statistically significant for two attributes versus one, for two versus three and for two versus four (all $p<0.01$), but not for three versus four. As in Experiment 1, we found no evidence of learning once the initial examples and practice trials were complete, despite hundreds of additional trials with feedback.

If a precision-bias trade-off does indeed govern the extent of bias across the price range then an individual-level analysis should also yield a positive relationship between precision and bias. This is a counterintuitive prediction, since it implies that the participants who were the best in terms of precision were the worst in terms of bias. To test this, we fitted the model specification of Model 3 in Table 2 to each of the 96 (24 x 4) experimental runs, then calculated the pairwise correlation between the estimated coefficients on the surplus and on the displayed price separately by number of attributes. Going from one to four attributes, the correlation coefficients were 0.35 ($p<0.1$), 0.67 ($p<0.01$), 0.24 (non-significant) and 0.62 ($p<0.01$). This positive correlation confirms that participants who were most precise were also most biased.

3.3.3 Discussion

Several minor changes in the experimental design may have contributed to the slightly better performance in the single-attribute condition compared with Experiment 1 and with multiple single-attribute S-ID tasks in Lunn and Somerville (2016). In particular, the product was constant and, because no condition required attributes to be traded off, the magnitude of the attribute and the product price were in perfect harmony throughout all practice and test presentations. In absolute terms, however, a surplus of 0.16 of the price range still equates to only seven discriminable levels of value, although there were approximately 26 perceptually discriminable levels of attribute magnitude. Additional attributes signalling the same surplus produced only a marginal improvement in precision. There was no evidence that multiple correlated attributes could be integrated efficiently, implying that where many attributes signal the same underlying quality, it has a significant but small impact on the ability to identify a surplus.

Experiment 3 implies that the psychological mechanisms behind surplus identification trade off precision in the centre of the price range against bias in the extremes. The results do not imply that the averaging of attribute magnitudes is not involved in judgements, but a process of averaging cannot explain the reversal in the relationship between the number of attributes and the bias across the price range, in comparison with Experiments 1 and 2.

4 General Discussion

Experimental control over attributes, prices and surpluses, within an incentive compatible design, is the essential characteristic of the S-ID task. This control allows a more systematic investigation of consumers' capabilities and associated cognitive constraints than is possible via the empirical examination of individual markets or behavioural biases. The technique permits the empirical separation of precision and bias, variation in which can be accurately estimated and related to variation in product characteristics. Hence, it is possible to use this novel task to investigate specific limitations on consumer choice and how they are likely to generalise across products and markets.

By way of demonstration, the three experiments presented here systematically investigated the impact on consumers' abilities to identify surpluses of increasing the number of product attributes from one to four. It is standard in consumer choice models not to limit the number of attributes that enter the utility function, or to condition the utility function on the number of attributes. Yet our results imply that the ability to identify a surplus is strongly constrained by the need to integrate information from multiple attributes to compare against price, especially where attribute magnitudes entail significant trade-offs. Surplus identification is imprecise and, furthermore,subject to a pattern of biases consistent with a trade-off between precision and bias.

These patterns of imprecision and bias have implications for current micro-economic thinking and invite further investigation. For instance, where choice does not involve probabilistic outcomes, e.g. between products under conditions of full information, it is generally characterised as "riskless choice". Yet the scale of imprecision implied by performance on the S-ID task with multiple attributes suggests that an alternative characterisation may be required. Consider a consumer who chooses between two products of equal consumption utility. The first has balanced attributes, but the second is good on one or more attributes and poor on one or more other attributes. Our results suggest that the consumer is likely to be less precise when assessing the surplus conferred by the second product and, furthermore, biased towards underestimating it. An intriguing possibility is that the imprecision and underestimation are related; they may reflect a form of built-in risk aversion. Either way, the consumer is inclined to judge the first product to have a higher surplus than the second, and is presumably more likely therefore to purchase it, yet is in fact indifferent between the two. Thus, underestimation of surplus in the face of trade-offs among attributes drives a wedge between revealed and true preferences. Analogous arguments can be made in relation to a consumer choosing between a simple product with one or two key attributes and a more complex one with additional features, or between a lower quality cheap product and a higher quality expensive one. In each case, versions of the S-ID task can be designed specifically to test whether the biases recorded when judging single products against prices generalise to choices among multiple products with prices.

In addition to systematic biases across the price range that depend on the number of attributes and bias in favour of balanced attributes, we recorded two other biases in the S-ID data (the dilution effect and a familiarity effect) first documented in choice experiments. This is in keeping with other recent empirical studies where context effects first demonstrated in choice experiments have also been located in tasks involving objective judgements (Trueblood et al., 2013; Lunn and Somerville, 2016), including the attraction, similarity and compromise effects, as well as reference dependence. As well as providing backing to our contention that common psychological mechanisms are engaged by the S-ID task and subjective choice tasks, these findings indicate a stronger link between the capacity constraints invoked by models of rational inattention (Sims, 2003; Woodford, 2014) and the context effects that motivate models of biased attribute weighting (Bordalo et al., 2013; Kőszegi and Szeidl, 2013; Bushong et al., 2015). The precision-bias trade-off in surplus identification that we highlight here further supports the argument that the psychological mechanisms involved in economic decisions are biased by neural adaptation to the likely range of inputs, designed to increase the acuity of relative comparisons Summerfield and Tsetsos (2015).

More generally, the S-ID task can be adapted in multiple ways to investigate other research questions of importance to the construction of more descriptively accurate models of consumer choice. Lunn and Somerville (2016) use the S-ID task to show that the ability to identify surpluses is unaffected by monotonic nonlinear preference functions, but that precision deteriorates when consumers face more complex nonlinear functions with turning points. In principle, the task can be adapted to test consumers' ability to adopt a preference function of almost any shape of interest. As well as extending the current findings beyond visual features, to categorical and numeric attributes, Lunn et al. (2016) deploy a version of the S-ID task with familiar products (houses and broadband packages) to show that neither the involvement of real products rather than a hyperproducts, nor the use of familiar attribute-price relationships, changes the basic pattern of imprecision and bias in surplus identification. The task has the potential to be further adapted to examine how precision and bias are affected by the number of products in the range, whether they have common attributes, the relative weightings of the attributes and characteristics of the distributions of these attributes, such as salience and relative ranges. The latter manipulation would permit direct tests of some features of specific consumer choice models cited above.

References

Agarwal, S., Chomsisengphet, S., Liu, C., and Souleles, N. S. (2015). Do consumers choose the right credit contracts? *Review of Corporate Finance Studies*, page cfv003.

Agarwal, S., Driscoll, J. C., Gabaix, X., and Laibson, D. (2009). The age of reason: Financial decisions over the life cycle and implications for regulation. *Brookings Papers on Economic Activity*, 2009(2):51–117.

Barber, B. M., Odean, T., and Zheng, L. (2005). Out of sight, out of mind: The effects of expenses on mutual fund flows. *The Journal of Business*, 78(6):2095–2120.

Barlow, H. (1961). Possible principles underlying the transformation of sensory messages. In Rosenblith, W., editor, *Sensory Communication*, chapter 13, pages 217–234. MIT Press.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843.

Bronnenberg, B. J., Dub, J.-P., Gentzkow, M., and Shapiro, J. M. (2015). Do pharmacists buy bayer? informed shoppers and the brand premium. *The Quarterly Journal of Economics*, 130(4):1669–1726.

Bushong, B., Rabin, M., and Schwartzstein, J. (2015). A model of relative thinking. *Unpublished manuscript, Harvard University, Cambridge, MA.*

Carlson, K. A., Meloy, M. G., and Russo, J. E. (2006). Leader-driven primacy: Using attribute order to affect consumer choice. *Journal of Consumer Research*, 32(4):513–518.

Chernev, A. (2005). Context effects without a context: attribute balance as a reason for choice. *Journal of Consumer Research*, 32(2):213–223.

Choi, J. J., Laibson, D., and Madrian, B. C. (2010). Why does the law of one price fail? an experiment on index mutual funds. *Review of Financial Studies*, 23(4):1405–1432.

Curry, D. J. and Faulds, D. J. (1986). Indexing product quality: Issues, theory, and results. *Journal of Consumer Research*, 13(1):134.

Garcıa-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision research*, 38(12):1861–1881.

Grubb, M. D. (2009). Selling to overconfident consumers. *The American Economic Review*, 99(5):1770–1807.

Hauser, J. (2011). A marketing science perspective on recognition-based heuristics (and the fast-and-frugal paradigm). *Judgment and Decision Making*, 6(5):396.

Heeley, D. and Buchanan-Smith, H. (1990). Recognition of stimulus orientation. *Vision Research*, 30(10):1429–1437.

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in Neuroscience*, 8(150).

Kőszegi, B. and Szeidl, A. (2013). A model of focusing in economic choice. *The Quarterly Journal of Economics*, 128(1):53–104.

Lambrecht, A. and Skiera, B. (2006). Paying too much and being happy about it: Existence, causes, and consequences of tariff-choice biases. *Journal of Marketing Research*, 43(2):212–223.

Laming, D. R. J. (1997). *The measurement of sensation*. Oxford University Press.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74(2):132–157.

Lee, J.-W. and Yates, J. F. (1992). How quantity judgment changes as the number of cues increases: An analytical framework and review. *Psychological Bulletin*, 112(2):363.

Lunn, P. D., Bohacek, M., and Somerville, J. J. (2016). Integration of visual, numeric and categoric information in surplus identification. In preparation.

Lunn, P. D. and Somerville, J. J. (2016). Experimental measures of consumers' ability to identify surplus with non-linear attribute returns. Working Paper 522, Economic and Social Research Institute.

Macmillan, N. A. and Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.

Meyvis, T. and Janiszewski, C. (2002). Consumers beliefs about product benefits: the effect of obviously irrelevant product information. *Journal of Consumer Research*, 28(4):618–635.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2):81.

Morgan, M. (1991). Hyperacuity. In Regan, M., editor, *Spatial vision*, pages 87–113. Macmillan London.

Morgan, M. (2005). The visual computation of 2-d area by human observers. *Vision research*, 45(19):2564–2570.

Morgan, M., Watamaniuk, S., and McKee, S. (2000). The use of an implicit standard for measuring discrimination thresholds. *Vision research*, 40(17):2341–2349.

Moscatelli, A., Mezzetti, M., and Lacquaniti, F. (2012). Modeling psychophysical data at the population-level: The generalized linear mixed model. *Journal of vision*, 12(11):26.

Phillips, F. and Todd, J. T. (2010). Texture discrimination based on global feature alignments. *Journal of vision*, 10(6):6–6.

Schkade, D. A. and Kahneman, D. (1998). Does living in california make people happy? a focusing illusion in judgments of life satisfaction. *Psychological Science*, 9(5):340–346.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, pages 99–118.

Simonson, I. and Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of marketing research*, 29(3):281.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Stevens, S. S. (1975). *Psychophysics.* Transaction Publishers.

Stewart, N., Brown, G. D., and Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review*, 112(4):881.

Summerfield, C. and Tsetsos, K. (2015). Do humans make good decisions? *Trends in cognitive sciences*, 19(1):27–34.

Troutman, C. M. and Shanteau, J. (1976). Do consumers evaluate products by adding or averaging attribute information? *Journal of Consumer Research*, 3(2):101–106.

Trueblood, J. S., Brown, S. D., Heathcote, A., and Busemeyer, J. R. (2013). Not just for consumers context effects are fundamental to decision making. *Psychological science*, 24(6):901–908.

Weaver, K., Garcia, S. M., and Schwarz, N. (2012). The presenter's paradox. *Journal of Consumer Research*, 39(3):445–460.

Wilson, C. M. (2010). Ordered search and equilibrium obfuscation. *International Journal of Industrial Organization*, 28(5):496–506.

Woodford, M. (2014). Stochastic choice: An optimizing neuroeconomic model. *The American Economic Review*, 104(5):495–500.

Woodward, S. E. and Hall, R. E. (2012). Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence. *The American Economic Review*, pages 3249–3276.

A Just Noticeable Difference (JND) and bias

The cumulative distribution function of a logistic distribution is standardly written as

$$F(x;\mu,s)=\frac{1}{1+e^{-(x-\mu)/s}} \tag{4}$$

where, at the mean, $x=\mu$, $F=0.5$, and where the distribution has variance $s^2\pi^2/3$. In the experiments, beginning with a baseline condition, e.g., a single-attribute experimental run of trials, $t=\{1,...,T\}$, where the dependent variable is whether the participant judged the surplus, Δ_t, to be positive, we fit a model of the form

$$Pr(\text{"}positive\text{"}) = \frac{1}{1 + e^{-(\gamma_0 + \gamma_1 \Delta_t)}} \tag{5}$$

Combining 5 with 4, note that $\mu = -\gamma_0 s$ and $s = 1/\gamma_1$. Thus, following estimation by maximum likelihood, we use the coefficients to estimate the surplus at which $Pr(\text{"}positive\text{"}) = 0.5$, giving

$$Bias = \frac{-\hat{\gamma_0}}{\hat{\gamma_1}} \tag{6}$$

For the JND we obtain the standard deviation from the estimated coefficients as follows

$$JND = \sqrt{\frac{\pi^2}{3\hat{\gamma_1}^2}} = \frac{\pi}{\hat{\gamma_1}\sqrt{3}} \tag{7}$$

Year	Number	Title/Author(s) ESRI Authors/Affiliates *Italicised*
2016		
	535	Evidence, Drivers and Sources of Distortions in the Distribution of Building Energy Ratings prior to and after Energy Efficient Retrofitting *Matthew Collins and John Curtis*
	534	The Impact of Free GP Care on the Utilisation of GP services in Ireland : An Evaluation of Different Approaches *Paul K Gorecki*
	533	An Examination of the Abandonment of Applications for Energy Efficiency Retrofit Grants in Ireland *Matthew Collins and John Curtis*
	532	An Examination of Energy Efficiency Retrofit Depth in Ireland *Matthew Collins and John Curtis*
	531	Use it or Lose it : Irish Evidence *Irene Mosca* and Robert E Wright
	530	Exporting under Financial Constraints: Firm-level Evidence from EU Countries Gavin Murphy and *Iulia Siedschlag*
	529	Fisheries Management for Different Angler Types *John Curtis and Benjamin Breen*
	528	Poorest made Poorer? Decomposing Income Losses at the Bottom of the Income Distribution During the Great Recession *Michael Savage*
	527	Profile of Second-level Students Exempt from Studying Irish *Emer Smyth* and *Merike Darmody*
	526	Modelling the Vietnamese Economy Pho Chi , *John FitzGerald*, Do Lam , Hoang Ha , Luong Huong, Tran Dung
	525	Attitudes to Irish as a School Subject among 13-year-olds *Emer Smyth* and *Merike Darmody*
	524	Attitudes of the Non-Catholic Population in Northern Ireland towards the Irish Language in Ireland *Merike Darmody*

For earlier Working Papers see http://www.esri.ie