

Do consumers take advantage of common pricing standards? An experimental investigation

Robert Sugden and Jiwei Zheng

Address (both authors): School of Economics, Centre for Behavioural and Experimental Social Science and Centre for Competition Policy, University of East Anglia, Norwich NR4 7TJ, UK; r.sugden@uea.ac.uk and j.zheng@uea.ac.uk.

Acknowledgements: We thank Alexia Gaudeul, Steffen Huck, Abhijit Ramalingam, Daniel Zizzo, an editor and three anonymous referees for valuable comments. Sugden's work was supported by the Economic and Social Research Council through the Network for Integrated Behavioural Science (grant reference ES/K002201/1) and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No. 670103.

Abstract: Gaudeul and Sugden have hypothesized that, when some but not all competing products are priced in a common standard, consumers who are liable to make errors in cross-standard price comparisons use decision rules that discriminate in favour of common-standard offers. Such behavior incentivizes sellers to use common standards. We report an experimental test of this hypothesis, using choice tasks similar to those represented in the Gaudeul–Sugden model. We found that offers priced in common standards were more likely to inspected but *less* likely to be chosen, and that subjects gained little benefit from common pricing standards that applied to some but not all offers. Most subjects used 'dominance editing' operations which eliminated transparently dominated offers, either as an initial shortlisting device or while offers were being sorted. Because these operations discriminate against common-standard offers, their use incentivizes sellers *not* to use common standards.

Keywords: shortlisting; common standard; dominance editing; consideration set

Do consumers take advantage of common pricing standards?

An experimental investigation

1. Introduction

One of the main mechanisms by which competitive markets are maintained is the ability and willingness of consumers to compare the offers of alternative suppliers and to buy at the cheapest price. As behavioural theories of individual choice have gained ground in economics, there has been increasing concern about the possibility that, by presenting offers with spurious complexity or ‘obfuscation’, firms can make price comparisons artificially difficult and so blunt price competition (e.g. Ellison, 2005; Gabaix and Laibson, 2006; Spiegler, 2006; Ellison and Ellison, 2009; Office of Fair Trading, 2011). Gaudeul and Sugden (2012) have pointed out that many forms of obfuscation work at the level of the market rather than at that of the individual firm. In these cases, the problem is not so much that the pricing schemes used by individual firms are unnecessarily complex in themselves, but rather that firms collectively do not quote prices in terms of *common standards* that would allow consumers to make simple price comparisons. For example, an often-cited form of obfuscation is the advertising of headline prices which exclude add-ons (such as delivery charges for products bought online, or baggage charges for air travel) that are over-priced or would customarily be included in quoted prices, or exclude excessive charges contingent on events which customers may fail to anticipate (such as overdraft fees on bank accounts). Except where add-ons are in fact unavoidable, breaking down services into separately-priced components expands consumer choice, and should not be a cause for concern in itself. The problem is that consumers tend to focus on headline prices when making overall price comparisons. Headline prices have more information content, the greater the degree to which there are market-wide conventions about what is included in headline prices and about how charges for add-ons are set. Obfuscation takes the form of the absence of such conventions, or the failure of firms to comply with them. Another way in which posted prices may not allow direct comparisons across products is through the use of non-standardized units of quantity or price. For example, different sellers of a product may use different package sizes and display per-package rather than per-unit prices, or use different conventions of measurement (such as different statistics for expressing interest rates), or offer idiosyncratic discounts (such as ‘three for the price of two’).

The persistence of these market-level forms of obfuscation is somewhat surprising, in view of the self-reinforcing properties of trading conventions in general. Economists have often argued that processes of competition spontaneously induce market-wide conventions about (for example) the medium of exchange, the specifications of complementary products, the location of shopping areas, and the dates and times at which trading takes place (Menger, 1892; Schotter, 1981; Sugden, 2004).

One might have expected a similar mechanism to lead to the emergence of common pricing standards. The intuition is that, since price comparisons are easier and more accurate when there are common standards, buyers who are drawing up shortlists of offers to consider in more detail will favour offers that are priced in common standards. Such a tendency would create incentives for sellers to conform to common standards as a way of gaining market share, and for retail platforms to enforce these standards.¹ Since competition is more intense when price comparisons are less noisy, these effects might lead to a virtuous circle: common standards could come to be signals of competitive prices, increasing the incentive for consumers to favour products priced in those standards. Gaudeul and Sugden (2012) present a model of this mechanism, and suggest that the absence of common standards in some real-world markets may be a sign that the spontaneous evolution of common standards has been obstructed by tacit collusion.

In this paper, we investigate whether buyers in fact discriminate in favour of common-standard offers. We report an experiment that elicited responses to price comparison tasks that were structured like those represented in Gaudeul and Sugden's model. Subjects were incentivized to search for the lowest price in a set of alternative offers. The experiment was designed to test two related hypotheses: that consumers benefit from the existence of common standards, even if all suppliers do not use them, and that consumers discriminate in favour offers that are priced in terms of such standards. If these hypotheses held for real-world consumers, the mechanism modelled by Gaudeul and Sugden would operate: there would be a self-reinforcing tendency for market competition to induce the emergence of common standards, and for that tendency to work to the benefit of consumers.

We compared behaviour in three types of task. In *all common standard* (AC) tasks, all offers were priced in a single common standard. In *part common standard* (PC) tasks, a subset of offers were priced in a common standard; each other offer had its own individuated standard. In *no common standard* (NC) tasks, every offer had an individuated standard. The results of our tests were negative. Although subjects were much better able to find low prices in AC tasks than in the other tasks, there was little evidence of their finding lower prices in PC tasks than in NC tasks. In PC tasks, with other relevant factors held constant, common-standard offers were more likely than other offers to be inspected, but *less* likely to be chosen.

To our initial surprise, however, we found patterns in our data that were suggestive of the effects of a decision rule which, as far as we know, has not previously been investigated in the context

¹ Ellison and Ellison (2009: 434-5) describe a case of such enforcement, which occurred during their study of Pricewatch, an internet search engine for the purchase of computer parts. Recognising that many sellers were adding excessive shipping fees at the check-out stage, and presumably aiming to protect the value of the price-comparison service that it was supplying, Pricewatch required all firms to offer a standard form of ground shipping with prescribed maximum charges.

of common pricing standards. When a shortlist of potentially choiceworthy offers is being constructed, this *dominance editing* rule eliminates any offer that is found to have a higher price than some other offer that is priced in the same standard. Thus, other things being equal, a firm's offer is more likely to be shortlisted, the *fewer* other offers are priced in the same standard. In principle, a consumer who uses this rule could (and rationally, ought to) recognize that the expected value of a shortlisted offer is higher, the more same-standard offers it has been found to dominate. But recognizing that requires sophisticated reasoning. If cross-standard price comparisons are subject to error, and if consumers treat all offers equally when making such comparisons, the dominance editing rule can have perverse effects on firms' market shares: consumers' attempts to take advantage of the simplifying properties of a common standard provide firms with incentives *not* to use that standard. Further analysis of our experimental data established that many subjects did indeed use dominance editing, and did not offset the resulting bias when making cross-standard comparisons. Thus, our findings provide evidence of the use of a decision rule which can obstruct the development of common pricing standards.

2. Shortlisting heuristics

For many years, psychologists and consumer researchers have been interested in two-stage *consider-then-choose* decision processes (e.g. Payne, 1976; Hauser and Wernerfelt, 1990; Hauser, 2010). Recently, economists have begun to investigate these processes too, and to consider the incentives that firms may have to tailor their offers to meet consumers' shortlisting criteria rather than to satisfy their actual preferences (e.g. Manzini and Mariotti, 2007; Eliaz, Richter and Rubinstein, 2011; Eliaz and Spiegel, 2011). The essential idea in these literatures is that, when a consumer faces a large choice set, he or she economizes on cognitive effort by using 'quick and dirty' heuristics to select a subset of options – the *shortlist* or *consideration set* – for further investigation. The final choice is then made from that subset using a procedure which tracks the consumers' preferences more accurately but is cognitively more costly. Shortlisting heuristics are understood as procedures that screen options relative to criteria that are easily observable, psychologically salient, and (in most models) positively correlated with the consumer's actual preferences. Examples of such heuristics include eliminating options that lack some pre-determined desirable quality (Tversky, 1972) or are transparently dominated by other options (Kahneman and Tversky, 1979), shortlisting options with recognized names (Goldstein and Gigerenzer, 2002), and shortlisting options that immediately engage attention (Eliaz and Spiegel, 2011).

Gaudeul and Sugden's (2012) model represents a different kind of two-stage decision-making. In this model, shortlisting is not an attempt to pick options that are expected to score highly on the scale of the consumer's true preferences. Instead, the aim is to construct the largest set of options that can easily and accurately be compared *with one another* on that scale. In other words,

the consideration set is constructed so that the second stage of decision-making can be carried out by using quick *and clean* processes. One significant feature of this form of shortlisting is that its use by consumers incentivizes (non-colluding) firms to converge on common standards.

In Gaudeul and Sugden's model, a consumer faces a *choice set* of n offers, from which she must choose one. Each offer i can be described by a pair (p_i, s_i) where p_i is the *final price* and s_i is the *standard* in which the price is expressed. The consumer's objective is to minimize the final price paid. She observes the standard of each offer, but is not directly informed about its final price. She is given information from which the final price can be reconstructed, but that reconstruction requires her to use cognitive processes that are liable to error. These processes are modelled by means of the assumption that, for each offer i , the consumer 'observes' a *price signal* $p_i + \varepsilon$, where ε is an i.i.d. random variable with zero mean and non-zero variance. The price signal is interpreted as the consumer's reconstruction of the final price of an offer whose actual price is p_i ; ε represents error. For simplicity, the distribution of ε is assumed to be independent both of the final price and of the standard in which the price is expressed. One way in which the consumer can form a preference between two offers is by comparing their price signals (i.e. by attempting to reconstruct their final prices). This is the *calculating* operation. However, she is also capable of forming preferences by using another, independent mental operation, the *ranking* operation. This operation generates an accurate ordinal ranking of every pair of prices p_i, p_j for which $s_i = s_j$, but cannot rank prices that are not expressed in the same standard.

If all offers are expressed in individuated standards, the consumer has to rely on the calculating operation; the best she can do is to choose the offer with the lowest price signal. But if the choice set contains at least two offers with a common standard, it is possible that the expected price she pays will be lower if instead she uses the *largest common standard (LCS) heuristic*. The first stage of this heuristic identifies the largest (or an equal-largest) set of offers such that all offers in this consideration set have a common standard. The second stage uses the ranking operation to find the lowest final price in the consideration set.

As a stylized example, consider a consumer facing a choice set of three offers. The final prices p_1, p_2 and p_3 are independent draws from a uniform distribution over $[1, 2]$. Offers 1 and 2 are expressed in a common standard; offer 3 is expressed in an individuated standard. (To fill in the story, suppose the consumer is buying bottled water, and wants to buy at the lowest price per litre. Suppliers quote their prices per bottle, and declare the capacities of their bottles. Firms 1 and 2 use the same size of bottle; firm 3 uses a different size. The consumer has difficulty with calculations that involve division.) If the consumer used the LCS heuristic, her expected payment would be 1.33 (the expected value of the lower of two randomly drawn prices). If she were able to use the calculating operation without error, her expected payment would be 1.25 (the expected value of the

lowest of three randomly drawn prices). But, depending on the variance of the error term, the expected value of the lowest of the three price signals can be greater than 1.33 (the upper limit of this expectation is 1.5).²

Gaudeul and Sugden's analysis is mainly concerned with how prices and standards evolve in a market in which some or all consumers use the LCS heuristic. But this heuristic is just one (particularly simple) way in which consumers can benefit by discriminating in favour of common-standard offers. Consider a consumer in the Gaudeul–Sugden model who faces a choice set of n offers, the final prices of which are independent draws from a common distribution. Suppose that two or more of these offers are priced in a single common standard, and that the rest are priced in individuated standards. The consumer initially knows only which offers are common-standard, and has to pick a shortlist of offers for which she will learn the price signals and (for common-standard offers) relative price rankings; she will then choose optimally from that shortlist. Assume that some constant cost, the same for both types of offer, is incurred for every shortlisted offer, or alternatively, that the number of shortlisted offers cannot be greater than some value m , where $1 < m < n$.

Since the consumer will receive unambiguously more information about each common-standard offer than about each individuated-standard one, it is immediately obvious that, for any given size of shortlist (greater than one), it is rational for her to include as many common-standard offers as possible. Equivalently, the more common-standard offers that she puts on a shortlist of given size, the lower is the expected final price paid. Thus, a rational consumer benefits from the presence of common-standard offers. The procedure by which she achieves these benefits discriminates in favour of such offers at the shortlisting stage.

Now consider how a rational consumer will choose from a shortlist in which some but not all offers are common-standard. Consider a shortlist containing m offers (where $m \geq 3$), of which m' (where $1 < m' < m$) share a common standard, the other offers being individuated-standard; as before, final prices are independent draws from a common distribution. To choose optimally, the consumer has to integrate the information contained in the m price signals with that contained in the ranking of the m' common-standard offers. This is a difficult mathematical problem for which we have been unable to find a general solution. However, it is clear that the optimal choice will always be *either* (a) the shortlisted offer with the lowest price signal *or* (b) the shortlisted common-standard offer with the lowest final price (as revealed by the ranking operation). If price signals are fully informative,

² As this example illustrates, the LCS heuristic has some similarities with modes of reasoning that generate the *decoy* (or *asymmetric dominance*) effect (Huber et al., 1982; Shafir et al., 1993). In the example, a consumer who uses the LCS heuristic recognizes that one of the common-standard offers dominates the other, and for that reason, chooses it from the set of three.

the optimal choice is always (a), and so the ex ante probability of choosing a common-standard offer is m'/m . If they are completely uninformative, the optimal choice is always (b), and so the ex ante probability of choosing a common-standard offer is 1. It seems intuitive that, the less informative price signals are, the more reliance the consumer will place on the ranking operation. We therefore conjecture that if price signals are less than fully informative, the ex ante probability of choosing a common-standard offer is greater than m'/m . In other words, when choosing from a shortlist, a rational consumer will discriminate in favour of common-standard offers.

Finally, consider the ex ante probability that a rational consumer will choose the *best* shortlisted offer (i.e. the shortlisted offer with the lowest final price). If she used only the calculating operation, this probability would be independent of whether that offer was common- or individuated-standard. But in fact the consumer receives additional information from the ranking operation. If the best shortlisted offer is common-standard, that offer is necessarily the best shortlisted *common-standard* offer, and so the ranking operation provides additional support for choosing it. Thus, the best shortlisted offer is more likely to be chosen if it is common-standard than if it is individuated-standard.

Since our concern is with the behaviour of consumers in real-world markets, we do not want to put too much weight on results derived from models of ideally rational decision-making. Nevertheless, the analysis we have presented provides some support for the hypotheses that our experiment was designed to test.

3. Experimental design: principles

Our experimental design used the same framework as the model described in Section 2. Subjects chose from sets of alternative offers, with an incentive to find the lowest final price. In most of our experimental tasks, final prices were not stated directly. Instead, subjects (who were not allowed to use calculating aids) were given information from which final prices could be calculated by addition and/or multiplication. For example, a final price of 10.07 ‘points’ (the experimental currency unit) might be expressed by means of the *price details* ‘10.60 points \times 95%’. We chose the parameters with the intention that calculating final prices would require some cognitive effort, and that subjects would be likely to make some errors in these calculations (or, alternatively, would choose to rely on approximations rather than exact calculations). Notice that the error mechanism in the experiment was a property of the subjects’ own reasoning, not a property of an external random mechanism. Our aim was to test how human decision makers choose in situations in which they might reasonably expect *their own decision-making processes* to be subject to error.

In each task, there were 24 offers, all with the same *offer type*. For example, the offers described by the price details ‘10.60 points \times 95%’, ‘9.50 points \times 95%’ and ‘12.20 points \times 86%’ all

have the type (which we will later denote ‘P*D’) in which the final price is described as ‘original price \times y%’, where y is a variable. Within a given task, two offers have a *common standard* if an accurate ordinal comparison of them can be made simply by ranking their original prices. For example, ‘10.60 points \times 95%’ and ‘9.50 points \times 95%’ have a common standard. An offer has an *individuated* standard if it cannot be compared in this way with any other offer in the task.

The experiment used four types of task. In the *benchmark* task, the final price of each offer was identical to the original price. All other tasks used offer types in which original and final prices were different. In *all common standard* (AC) tasks, all 24 offers had a common standard. In *no common standard* (NC) tasks, all 24 offers had individuated standards. In *part common standard* (PC) tasks, eight of the 24 offers had a single common standard; the other 16 offers had individuated standards. The PC tasks are the tasks most directly related to the real-world target of our work, namely markets in which some but not all products are priced in common standards. However, by comparing responses to NC and PC tasks we were able to assess the incremental effect of common standards on subjects’ ability to make accurate price comparisons. The AC tasks allowed us to test whether subjects were able to recognize that common-standard offers could be compared simply by ranking their original prices. Since recognizing this is a precondition for understanding the value of discriminating in favour of common-standard offers in PC tasks, experience of AC tasks should be expected to facilitate such discrimination. By using equal numbers of NC, AC and PC tasks, we tried to avoid steering subjects either towards or away from using the ranking operation.

In all tasks, the distribution of final prices was the same. Final prices were uncorrelated with standards, and so the fact that a particular offer was or was not priced using a common standard contained no information about its final price. This provided a neutral setting for investigating how (if at all) individuals made use of common standards when choosing between offers. The experimental interface between subject and task was designed to allow us to track significant aspects of subjects’ reasoning, and thus to look for the fingerprints of heuristics that subjects might be using. At the start of any task, offers were presented in a way that allowed subjects to identify which, if any, used a common standard, but subjects were required to make specific mouse clicks to reveal further information about specific offers. These mechanisms enabled us to track which offers were inspected in which order.

The experiment was designed on the working assumption that subjects are less able to make accurate price comparisons across standards than within a common standard. The following hypothesis states this assumption in a testable form:

Hypothesis 1: The average final price of chosen offers is lower in AC tasks than in NC tasks.

As explained in Section 1, the experiment was designed to test two general hypotheses: that consumers benefit from the existence of common standards, even if those standards are not in universal use, and that when some but not all offers are priced in a common standard, consumers favour common-standard offers. In the context of our design, the first of these hypotheses can be expressed as:

Hypothesis 2: The average final price of chosen offers is lower in PC tasks than in NC tasks.

The idea that common-standard offers are favoured can be expressed in several different ways. We will be concerned with three kinds of favouring, expressed in the following hypotheses:

Hypothesis 3: In PC tasks, the proportion of common-standard offers that are inspected is greater than the corresponding proportion of individuated-standard offers.

Hypothesis 4: In PC tasks, the probability that the chosen offer is common-standard is greater than the relative frequency of common-standard offers in the set of inspected offers.

Hypothesis 5: In PC tasks, the probability that the best inspected offer (i.e. the inspected offer with the lowest final price) is chosen is higher if the best inspected offer is common-standard than if it is individuated-standard.

Hypothesis 3 proposes that, other things being equal, individuals are more likely to inspect offers that use common standards. Recall that, in our experiment, subjects were given no offer-specific information about the final price of an offer until they inspected it. Thus, with respect to final prices, each subject's set of inspected offers was a random sample of the set of all 24 offers, irrespective of whether common-standard offers were favoured or disfavoured at the inspection stage. Hypotheses 4 and 5 propose that, after controlling for differential rates of inspection, individuals favour common-standard offers when choosing among inspected offers. Hypothesis 4 is framed in terms of common-standard offers in general, and so aggregates across common-standard offers with different final prices, while Hypothesis 5 refers only to the best available inspected offer.

As explained in Section 2, Hypotheses 2, 3 and 5 (and, we conjecture, Hypothesis 4) are implications of the Gaudeul-Sugden model under the assumption of rational choice. However, these hypotheses are framed in terms of benchmarks that are significant for the functioning of markets. Whether or not consumers are ideally rational, it is important to know whether they benefit from common standards (Hypothesis 2), whether common-standard offers in general are more likely to be chosen than individuated-standard ones (as implied by Hypotheses 3 and 4), and whether a firm that sets a final price below those of its competitors can gain market share by pricing in a common standard (as implied by Hypotheses 3 and 5).

4. Experimental design: details

The experiment had a within-subject design. Each subject completed ten different tasks presented in randomized order on a computer screen. In each task, a subject was given an ‘endowment’, and was required to buy one unit of a notional good by accepting one of 24 alternative price ‘offers’. All final prices were lower than the endowment. Endowments and prices were expressed in ‘payment points’. In each task, the subject was credited with payment points equal to the endowment minus the final price of the chosen offer. At the end of the experiment the computer randomly picked one of the ten tasks; the subject’s earnings from the experiment were equal to her payment points from that task, converted into UK pounds using an ‘exchange rate’ that was specific to the task. Thus, in each task, subjects had an incentive to choose the offer with the lowest final price.

In every task, the money value of the endowment (converted at the task-specific exchange rate) was £32. The set of final prices, expressed in UK pounds, was the same in all tasks, except for rounding; these unrounded prices ranged from £20.270 to £28.895, in increments of £0.375. Keeping the set of final prices constant was important in allowing controlled hypothesis tests. However, it was also important that subjects did not recognize this feature of the design and use it to simplify their decision problems (for example, by remembering the value of the lowest price). Thus, although subjects were told at the start of the experiment that the endowment would always have a converted value of £32, they were *not* told that the distribution of converted final prices was the same in all tasks. Nor were they told anything about these distributions other than that, in every task, the highest price was always lower than the endowment. The exchange rate was generated randomly, independently for each subject and for each task. It was expressed in the form ‘ x points = £1’, where x was a round number in the interval $10 \leq x \leq 100$. The price details of the offers in a task were then constructed so as to be consistent with the relevant exchange rate. For example, an offer with a final price of £20.27 and the offer type P*D might appear in one task with an exchange rate of 12 points = £1 and an endowment of 384 points, and be described as ‘original price = 676 points; final price = original price \times 36%’. The same offer might appear in another task with an exchange rate of 43 points = £1, an endowment of 1376 points, and the description ‘original price = 1063 points; final price = original price \times 82%’.

In any given task, all offers were expressed in terms of the same offer type. The following four offer types were used:

(1) *Type P*: Price details are described as a single price such as:

$$\text{Final price} = x \text{ points}$$

where x is a positive round number. This offer type was used only in the benchmark task. In that task, offers were differentiated only by their original prices. In tasks with this offer type, we will say that all offers have the same *price structure*.

(2) *Type P*D*: Price details are described as an original price with a percentage discount:

$$\text{Original price} = x \text{ points}; \text{Final price} = \text{Original price} \times y\%$$

where x is a positive round number and y is a round number in the interval $0 < y < 100$. In tasks with this offer type, we will say that offers have the same price structure if and only if they have the same value of y .

(3) *Type P*D+A*: Price details are described as an original price with a percentage discount plus an add-on:

$$\text{Original price} = x \text{ points}; \text{Final price} = \text{Original price} \times y\% + z \text{ points}$$

where x and z are positive round numbers and y is a round number in the interval $0 < y < 100$. In tasks with this offer type, offers have the same price structure if and only if they have the same values of both y and z .

(4) *Type P*D1*D2*: Price details are described as an original price with two percentage discounts:

$$\text{Original price} = x \text{ points}; \text{Final price} = \text{Original price} \times y_1\% \times y_2\%$$

where x is a positive round number and y_1 and y_2 are round numbers in the intervals $0 < y_1 < 100$ and $0 < y_2 < 100$. In tasks with this offer type, offers have the same price structure if and only if they have the same values of both y_1 and y_2 .

Whenever two offers in the same task have the same price structure, the ranking of their final prices is necessarily the same as the ranking of their original prices. Thus, if two or more offers in a task have the same price structure, they have a common standard as defined in Section 2; if some price structure is unique to a specific offer, that offer has an individuated standard.

As explained in Section 3, the experiment used four types of task – benchmark, AC, NC and PC. There was one benchmark task, using the offer type P. Nine further tasks were constructed by crossing the AC, NC and PC task types with the P*D, P*D+A and P*D1*D2 offer types. (Since offer type P necessarily implies that all offers have a common standard, this design is fully factorial.) In PC tasks, which eight of the final prices £20.270, ..., £28.895 were assigned to the common standard was determined randomly, independently for each subject. Thus, the standard in which an offer was expressed provided no information about the final price. The ten tasks used in the experiment are summarized in Table 1.

[Table 1 near here]

The four offer types were chosen to cover a range of levels of complexity. The P type is the simplest that is possible in our experimental set-up. The P*D type is unambiguously less complex than the P*D+A and P*D1*D2 types. By using tasks with different degrees of complexity, we made it more likely that each subject would face some tasks in which the problem of working out final prices was perceived as difficult but manageable. This design feature also allows us to investigate

whether the decision rules that subjects use are affected by the complexity of the offer type. Subjects were not allowed to use calculating aids or to make calculations on paper. This rule is stated in the instructions routinely used in our lab. As an additional precaution, on entering a session, subjects were required to put all bags, coats, phones, paper and pens on a table at the front of the lab. The settings on the computers in the lab did not allow access to installed calculators.

Each task was presented in the form of two screen pages – the *market page* and the *shopping basket page*. Screen shots of these pages are shown in Appendix B. (All Appendices are in the Electronic Companion.) At the start of the task, the subject saw the market page. The subject's endowment (in points) and the exchange rate between points and UK pounds were displayed at the top of the page. One panel of the page displayed 24 coloured boxes, corresponding with the 24 offers, which were described as the offers of 24 different 'firms'. The boxes were labelled 'Offer A', ..., 'Offer X'. The allocation of actual offers to boxes was randomized. Boxes had the same colour if and only if the corresponding offers had the same price structure. In all other respects, colours were randomized. No details of the offers apart from the colours and the letter codes were visible.

In the market page, a subject was able to do four types of action. (1) She could single-click on an offer, in which case the offer was highlighted and a message such as 'price structure 1' would immediately appear at the bottom of the page.³ (2) She could click more than once on an offer, in which case the relevant price details would appear at the bottom of the screen after a three-second delay. (Actions (1) and (2) could be done only for one offer at a time: if the subject clicked on a new offer, information about the price structure and/or price detail of the previous offer disappeared.) (3) Following action (1) or (2), the subject could click a 'Move into the shopping basket' button, in which case a correspondingly coloured and labelled box immediately appeared in a 'shopping basket' panel of the market page. The shopping basket had a maximum capacity of nine offers; subject to this limit, the subject could move as many offers to the basket as she wished without any time delays. (4) At any time, the subject could click a 'View the shopping basket' button, in which case she would move to the shopping basket page after a three-second delay.

In the shopping basket page, the subject could immediately see the price details of all the offers that had been moved into the basket (and had not yet been moved out), displayed in the middle of the screen. Thus, price comparisons (particularly for offers with complex price structures) could be made much more easily in the shopping basket page than in the market page, where the subject could view the price details of only one offer at a time. In this page, the subject was able to do three

³ Since the numbering of price structures was arbitrary, and since the colours of the boxes showed which if any offers had the same price structure, this message had no real information content. It served as an aid for any subjects who had difficulty distinguishing colours.

types of action, none of which involved any time delay. (1) She could move any offer(s) out of the basket, thus creating space to add new offers from the market page. (2) She could go back to the market page by clicking a ‘Continue shopping’ button. (3) She could make a final decision by clicking a ‘Buy the good from firm —’ button (where — was the letter identifying the chosen offer).

Because a final choice could be made only on the shopping basket page, because offers could be compared more easily on that page, and because the only way to tag an offer as having special status was to move it to that page, we expected subjects to use the shopping basket to store offers that they perceived as potentially choiceworthy. If the shopping basket was used in this way, the evolution of its contents over the course of a task could be interpreted as the progressive building up and refinement of a shortlist, and so would be informative about subjects’ reasoning.

The three-second time delays served two purposes. First, the existence of short time delays replicates the fact that in real-world internet shopping there is always some time delay while opening and negotiating a website to find the price of a specific product, or going to a shopping basket page for checking out. Second, the delay mechanism was intended to encourage subjects to use the shopping basket as a shortlist. If comparisons were made in the market page, there was a three-second delay for every offer inspected, but if offers were moved to the shopping basket before being compared, there was only one such delay for every batch of (up to nine) offers moved together. However, no time constraints were imposed. Subjects were free to spend as long as they liked on any task.

Individual subjects were paid and were free to leave the experiment after completing all their tasks; they did not have to wait until others had finished. Thus, it is reasonable to assume that time spent inspecting and comparing offers had an opportunity cost for our subjects. In imposing search costs in the form of losses of time rather than of money, our design reflects most consumers’ everyday experience of making price comparisons – for example, by visiting competing stores in a shopping mall, or visiting the websites of competing suppliers.

Before starting the formal experiment, the experimenter read out the experimental instructions and activated all subjects’ computer screens to let them do a practice task. The practice task was similar to the PC tasks but with a different offer type. (This had the form: *Original price = x points; Final price = Original price × y%* with $y > 100$.) Subjects went through the practice task step by step, following the experimenter’s instructions. After finishing the practice task, subjects were given a questionnaire to check their understanding of the experiment. The formal experiment began only after these questions had been answered correctly. The instructions and questionnaire are reproduced in Appendix A.

5. Results: tests of Hypotheses 1–5

The experiment was conducted at the Centre for Behavioural and Experimental Social Science Laboratory at the University of East Anglia in the summer of 2013. It was implemented using z-Tree (Fischbacher, 2007). 171 subjects were recruited using a campus-wide online system. Most of the subjects were students from a wide range of academic disciplines, with ages ranging from 18 to 65. The experiment lasted approximately 65 minutes with an average payment of £10.76 per person. Payments ranged from £3.10 to £11.73.

Table 2 reports means, medians and standard deviations of the time used and of the variable *ranked payment*. This variable measures the final prices of the offers chosen by subjects in the relevant task. It uses a linear transformation of the final price scale; the value 1 denotes the lowest of the 24 final prices (£20.27 in all tasks) and the value 24 denotes the highest of these prices (£28.90). Table 2 also reports, for each task, the percentage of subjects who chose the *objectively optimal* offer (i.e. the offer with rank 1).

[Table 2 near here]

In the benchmark task, the low mean and median values of ranked payment (1.88 and 1 respectively) and the high proportion of objectively optimal choices (89.5 per cent) show that most subjects had little difficulty in negotiating the market and shopping basket pages and were sufficiently patient to search for the lowest price, despite the time delays built into the experimental interface. The time used in this task (mean 158 seconds, median 125 seconds) provides a useful baseline from which to measure the extra time used in making the more difficult price comparisons required by other tasks.

Time used and ranked payments in the AC tasks are only slightly higher than in the benchmark task, and the proportions of optimal choices are only slightly lower. Given the complexity of the arithmetic operations required to calculate final prices in the AC tasks, it is natural to infer that most subjects realized that, when offers have the same price structure, such calculations were not necessary; they simply compared original prices. The data from NC tasks give further support to this inference. It is immediately obvious from Table 2 that subjects found NC tasks much more difficult and time-consuming than AC tasks. For each of the three offer types, mean and median values of time used and ranked payment are much higher in NC tasks than in AC tasks, and the proportion of optimal choices is much lower. In all cases, these differences are highly significant ($p < 0.001$), confirming Hypothesis 1.⁴ Since the only difference between AC and NC tasks is the

⁴ Throughout the paper, unless otherwise stated, all within-subject tests are Wilcoxon signed rank tests, while for between-subjects comparisons we use Mann-Whitney tests. All p -values are two-sided. We say that test results are ‘significant’ and report p values when $p < 0.10$; if $p \geq 0.10$, we say that results are ‘not significant’.

presence or absence of a common standard, it is clear that a common standard, if used by all offers, makes decision-making easier and less subject to error.

In light of the evidence reported in the previous paragraph, it seems clear that most of our subjects had the ability to use the ranking operation, and hence to use the LCS heuristic. In a PC task, a subject who can use this heuristic has the capacity to find the offer with the lowest final price in the set of common-standard offers. The ranked payment for that offer (which has an expected value of 2.68 in all tasks) provides a measure of what a relatively unsophisticated subject might realistically achieve in a PC task by taking account of ranking information. (It sets an upper bound to the ranked payment that would be achieved by a subject who was ideally rational in the sense discussed in Section 2.) Actual ranked payments in NC tasks provide a comparable measure of what subjects were able to achieve without using ranking information. For each of the three offer types, mean and median ranked payments in NC tasks are higher than 2.68 (see Table 3). The distributions of ranked payments in NC tasks are significantly different from those that would be induced by using the LCS heuristic ($p < 0.001$ in all three cases, using Mann-Whitney tests). The implication is that a subject would do better in PC tasks by using the LCS heuristic than by ignoring the common/individuated distinction. Comparing time used in AC and NC tasks, and given that in a PC task the LCS heuristic requires only eight offers to be inspected and ranked, it seems likely that a typical subject would also save at least three minutes in each PC task by using the LCS heuristic rather than treating these problems like NC tasks.

However, the data in Table 2 provide little support for Hypothesis 2. For each offer type, mean and median values of ranked payment are similar in PC and NC tasks; there is a significant difference only for the P*D offer type (where chosen offers have lower prices in PC tasks; $p = 0.056$). Other comparisons between PC and NC tasks show a similar picture. Time used is not significantly different between PC and NC tasks for any offer type. The proportion of optimal choices is significantly different only for the P*D offer type (where it is higher for PC tasks; $p = 0.029$).

Table 3 reports the numbers of offers inspected in each task. An offer is defined to have been *inspected* by a subject if and only if the subject has seen its price details, either by double-clicking it on the market page or by moving it to, and viewing it in, the shopping basket page. In PC tasks, the proportion of offers inspected was significantly higher for common-standard offers than for individuated-standard offers ($p < 0.001$ for each of the three offer types), consistently with Hypothesis 3. In every task, however, most subjects inspected most (and often all) of the 24 offers, contrary to the implications of the LCS heuristic.⁵

⁵ For the P*D, P*D+A and P*D1*D2 offer types respectively, only 1, 1 and 2 of the 171 subjects inspected no individuated-standard offers, and only 7, 10 and 12 inspected fewer than four such offers.

[Table 3 near here]

Table 4 reports, for each PC task, the proportion of subjects who chose common-standard offers in each PC task. This is compared with the proportion of all inspected offers (summed over all subjects) that are common-standard. Contrary to Hypothesis 4, the former proportion is *less than* the latter in all three tasks; this difference is significant for the P*D+A and P*D1*D2 offer types ($p < 0.001$ and $p = 0.011$ respectively in two-sided binomial tests) but not for the P*D offer type. Notice also that, in each PC task, the proportion of subjects who chose common-standard offers is less than one-third (i.e. the proportion of *all* offers that are common-standard). Thus, despite the fact that common-standard offers were more likely than individuated-standard offers to be inspected, they were less likely to be chosen.

[Table 4 near here]

Table 5 reports the frequency with which the best inspected offer was chosen in each PC task, conditional on whether that offer was common-standard or individuated-standard. In the P*D task there is no significant difference between the two conditional frequencies. In the P*D+A task, this frequency is significantly greater when the best inspected offer is individuated-standard ($p = 0.050$), contrary to Hypothesis 5. In the P*D1*D2 task, the difference is consistent with Hypothesis 5 and highly significant ($p < 0.001$). In Section 7, we will suggest a possible explanation for this conflicting evidence.

[Table 5 near here]

Our conclusions so far can be summarized in the following results:

Result 1: Relative to the case in which all offers have individuated standards, buyers are better able to find low prices if all offers have a single common standard. But this effect is weak or non-existent if some but not all offers have a common standard.

Result 2: When some but not all offers have a common standard, common-standard offers are more likely to be inspected than are individuated-standard offers.

Result 3: When some but not all offers have a common standard, inspected common-standard offers are less likely to be chosen than are inspected individuated-standard offers.

Result 4: The effect described by Result 3 is stronger than that described by Result 2: when some but not all offers have a common standard, common-standard offers are less likely to be chosen than are individuated-standard offers.

Result 5: The evidence does not support a firm conclusion about whether, when some but not all offers have a common standard, the best inspected offer is more likely to be chosen if it is common-standard than if it is individuated-standard.

6. Dominance editing

The results reported in the previous Section pose a puzzle. It is clear that, when some but not all offers were priced in a common standard, subjects discriminated between those offers that were priced in this way and those that were not. This discrimination had the effect that common-standard offers were more likely to be inspected but less likely to be chosen. Overall, subjects gained little if anything (either in money or time) from the existence of common standards that allowed accurate comparisons among a subset of offers, despite the fact that comparisons between other offers were subject to considerable error. The puzzle is to identify the property (or properties) of subjects' decision processes that induced these distinctive patterns. The methodology for this part of our investigation was inductive, guided by conjectures formed after reflecting on the findings reported in Section 5.

We should make clear that our concern is with properties of decision processes that are specifically related to common standards and so are potential explanations of the regularities we have found. Since almost all subjects inspected some individuated-standard offers in PC tasks and most subjects inspected many, it is clear that subjects often needed to make cross-standard price comparisons in those tasks. There are many ways in which such comparisons might be made. For example, a subject might try to reconstruct the final price of each offer, considered separately, and then select the offer that is judged to have the lowest final price. Or she might use this calculating operation in conjunction with a stopping rule, based on an aspiration level (Simon, 1956) or a reservation price (Diamond, 1971; Salop and Stiglitz, 1977). Or she might use some heuristic which avoids trade-offs by treating similar magnitudes as equal (Tversky, 1969; Rubinstein, 1988) or by prioritizing particular dimensions of price information (Brandstätter et al., 2006). For our purposes, however, it is not necessary to know exactly how cross-standard comparisons were made. Our experiment was not designed to investigate this issue.

It is useful to distinguish between three components of the decision process – *inspection*, *sorting*, and *choice*. Through inspection, a subject arrives at a subset of offers for which price details are known. Recall that, in terms of final prices, this subset is a random sample of the whole set of 24 offers. By sorting among inspected offers, the subject arrives at a *final basket* – the contents of the shopping basket immediately before a final choice is made. The subject then chooses one offer from the final basket. In testing Hypotheses 2, 3 and 4, we have already isolated the effects of inspection

in PC tasks. For further investigation of decision processes, it is useful to isolate the effects of sorting in these tasks.

Table 6 reports the proportions of inspected common- and individuated-standard offers that ended up in the final basket. This reveals a very strong and robust regularity: in all three PC tasks, inspected common-standard offers are under-represented in the final basket. On average, an inspected offer is about twice as likely to be in the final basket if it is individuated-standard than if it is common-standard. For all three tasks, this difference is overwhelmingly significant ($p < 0.001$). Table 7 reports the frequency with which the *best* inspected offer ends up in the final basket. At this level there is no consistent asymmetry: there is discrimination against common-standard offers in the P*D+A task ($p = 0.015$) but in favour of such offers in the P*D1*D2 task ($p < 0.001$); in the P*D task there is no significant discrimination either way. The overall implication is that the sorting process discriminates against common-standard offers *as a class*, and does so by being more likely to reject high-price common-standard offers than to reject equivalent individuated-standard offers.

[Tables 6 and 7 near here]

Table 8 isolates the final component of the decision process by reporting the frequencies with which the offer that is best *in the final basket* is chosen. For each of the three tasks, the proportion of cases in which the best offer in the basket is chosen does not differ significantly according to whether that offer is common- or individuated-standard. In other words, there is no evidence of discrimination at the choice stage, either in favour of common-standard offers or against them.

[Table 8 near here]

These findings are suggestive of an operation similar to one of the editing operations that are applied to lotteries in prospect theory (Kahneman and Tversky, 1979: 274–275).⁶ In the context of our experiment, one inspected offer *dominates* another if they are both common-standard and if the original price of the former is less than that of the latter. In such a case, it is easy to recognize that the dominated offer will not be the best inspected offer. A subject who uses the *dominance editing* operation eliminates offers from consideration as soon as she becomes aware that they are dominated. If the shopping basket is used to store offers that are currently considered potentially choiceworthy, the effect of dominance editing will be to exclude or remove from the basket any common-standard offer that is inferior to the best such offer that has so far been inspected. This mechanism would generate an overall tendency to discriminate against common-standard offers during sorting, while exempting the *best* inspected common-standard offer from this discrimination.

⁶ In the original version of prospect theory, the method by which prospects are evaluated can sometimes allow stochastically dominated prospects to have higher values than the prospects that dominate them. Kahneman and Tversky postulate an editing operation which eliminates transparently dominated alternatives before any prospects are evaluated.

Of course, dominance editing is not irrational in itself; in principle, it can be part of an optimal decision rule. For example, a subject who makes fully accurate cross-standard price comparisons will end up choosing the best inspected offer whether or not she uses dominance editing. A second possibility is that of a subject who knows that her cross-standard price comparisons are subject to error and makes rational use of this knowledge. Such a subject might use dominance editing in the sorting process, but offset the resulting bias by discriminating in favour of common-standard offers when comparing them with individuated-standard offers. (Intuitively, the subject would take account of the information content of the fact that, at any given moment, the surviving common-standard offer is the best of the common-standard offers that have so far been inspected.) This kind of discrimination could occur either in the sorting process or at the final choice stage (or both). However, neither of these possibilities is consistent with our findings. The evidence from NC tasks shows that subjects did not make accurate cross-standard comparisons. Had subjects made rational use of the information available to them in PC tasks, the frequency with which common-standard offers were chosen would have been greater than the frequency with which they were inspected, but we found the opposite. These considerations suggest that dominance editing, combined with insufficient offsetting of the biases it induces, may have been responsible for the relatively high final prices paid in PC tasks.

One simple and salient form of dominance editing is to begin by inspecting only common-standard offers, eliminating those that are dominated before considering individuated-standard offers. As a first step in investigating whether our subjects used dominance editing, we therefore looked at the order in which offers were inspected in PC tasks. Table 9 shows, for each PC task and for each $m = 0, \dots, 8$, the number of subjects whose first eight distinct inspections included exactly m common-standard offers.⁷ For subjects who did not differentiate between common-standard and individuated-standard offers and who inspected eight or more offers, the distribution of m would be binomial, with a mean of 2.67; values of 7 and 8 would lie outside the 99 per cent confidence interval.⁸ Clearly, the hypothesis that inspections are randomly distributed must be rejected. The observed distributions are very obviously bimodal, suggesting that a majority of subjects began PC tasks by inspecting only, or almost only, common-standard offers, while a large minority began by inspecting offers in random order.

[Table 9 near here]

⁷ Very small numbers of subjects (6, 9 and 7 in the P*D, P*D+A and P*D1*D2 tasks respectively) inspected fewer than 8 offers. These subjects are classified according to the total number of common-standard offers they inspected.

⁸ Since the positions of the 24 offers on the market page were randomized independently for each subject, any inspection strategy that ignores the distinction between common and individuated standards necessarily inspects offers in random order.

This finding raised the possibility that there might have been two importantly different modes of behaviour in our experiment. To explore this possibility, we made a binary classification of subjects, separately for each PC task. In each such task, subjects were classified as *Group 1* if they inspected seven or eight common-standard offers in their first eight distinct inspections (or if they made only seven inspections, all common-standard, in the whole task), and as *Group 2* otherwise. We then investigated whether our main results held for each of these groups separately. In Appendix C we disaggregate the data on PC tasks in Tables 2–8 between the two groups. Unsurprisingly, given how the groups are defined, their inspection behaviour was different. In all three PC tasks, the frequencies with which offers were inspected in Group 1 were significantly greater for common-standard offers than for individuated-standard ones, while these frequencies were approximately equal in Group 2. But the sorting and choice processes of the two groups revealed similar qualitative patterns. The sorting processes of both groups showed strong and highly significant (post-inspection) discrimination against common-standard offers, but showed no consistent pattern of discrimination for or against the *best* inspected common-standard offer. For neither group was there any evidence of discrimination in favour of common-standard offers when choosing from the final basket.

Further evidence that the distinction between Groups 1 and 2 is not critical for an explanation of the overall bias against common-standard offers is provided by a regression analysis, reported in full in Appendix D. The following is a brief summary of that analysis, as applied to PC tasks. In an attempt to explain subject-level observations of ranked payments (which can be interpreted as an inverse measure of quality of decision) in PC tasks, we estimated an ordered probit model with random effects. The independent variables were: the offer type, the number of offers inspected, the time spent on the task, the position of the task in the series of ten tasks faced by the subject, the subject's age, the subject's gender, whether the subject's main field of study was related to mathematics or economics, whether the subject was classified as Group 1 or Group 2, and the subject's classification in terms of 'one-click' and 'two-click' inspections (see footnote 9 in Section 7 below). We found significant effects ($p < 0.05$) for offer type (decision quality was highest for the P*D offer type and lowest for P*D1*D2), number of offers inspected (larger numbers of inspections were associated with better decisions), time spent (decision quality improved with time spent), position (decision quality improved over the course of the experiment),⁹ and exposure to mathematics or economics (decision quality was higher for subjects with such exposure). None of the other variables was significant. The significant effects are not particularly informative about the precise decision processes used by subjects, but they are consistent with a general picture of subjects trying to

⁹ The effect of position was not consistent across task types. The effect of position (measured from 1 to 10, with 1 representing the first task faced) on decision quality was positive in PC tasks ($p = 0.015$), negative in NC tasks ($p = 0.069$), and not significant in the (very easy) AC tasks.

find the lowest final price in the set of inspected offers, applying fallible mental arithmetic to problems of varying difficulty.

7. The sorting process

To gain further insights into the sorting process, we looked at the evolution of shopping basket contents over the course of each PC task. To allow a simple overview of the relevant data, we need an interpersonally comparable unit for measuring a subject's progress through a task. We have found that the most informative way to do this is to use a measure which lines up the end of each task for all subjects. Since progression through a task is determined by mouse clicks, we use a measure of *clicking time*. We define percentiles of clicking time separately for each subject and each task. If a given subject made a total of N clicks during a task, her t th click is assigned to the $(100t/N)$ th percentile of clicking time. Averaging across subjects, the value of N was fairly stable across tasks (ranging from 57.6 to 68.6), but the standard deviation was always quite large (ranging from 15.6 to 28.8).¹⁰ In the graphs that we will present, each unit of observation corresponds with a 10 percentage point interval of clicking time.

For any given offer in any given task, for any given group of subjects, and for each interval of clicking time, we define the *in-basket probability* as the probability that, for a randomly selected subject from that group and for a randomly selected moment in that interval, the relevant offer is in the subject's shopping basket. Formally, for any given subject, we define *moments* $t = 1, \dots, N$, each corresponding with one click. For each moment t , for any given offer, we define the offer's *basket status* B_t as 1 if the offer is in the shopping basket immediately after the subject's t th click and 0 if it is not. For each interval of clicking time, the probability that the offer is in the subject's shopping basket is defined as the mean value of B_t , averaged over all moments in that interval. The mean of these probabilities, averaged over all subjects for a given interval of clicking time, is the relevant in-basket probability. Figure 1 presents graphs of the evolution of the in-basket probabilities of different offers in the P*D task (task 8), for Group 1 and Group 2 subjects separately. We focus on just one PC task in the interests of brevity. Corresponding graphs for the P*D+A and P*D1*D2

¹⁰ One source of variation was a sharp division between *one-click* subjects who moved offers to the shopping basket before inspecting price details and *two-click* subjects who inspected price details in the market page before deciding whether to move offers to the basket. In relation to the evolution of basket contents, the main difference between these two inspection strategies is that the one-click method requires every inspected offer to be in the basket for some interval of clicking time, while the two-click method allows unattractive offers to be eliminated without ever appearing in the basket. However, the qualitative patterns we will discuss were common to one-click and two-click inspection strategies. There was a weak tendency (statistically significant only for the P*D+A task) for subjects who used the one-click strategy to be more likely to belong to Group 1. The regression analysis (reported in Appendix D) found no significant difference between the final prices of the offers chosen by one-click and two-click subjects.

tasks are presented in Appendix E. The broad qualitative patterns we will identify are common to all three PC tasks.

[Figure 1 near here]

We begin by looking at the evolution of in-basket probabilities for common-standard offers and for Group 1 subjects. The relevant graphs are shown in Figure 1a. We classify offers according to their final prices. Since the best common-standard offer has the special property that it can never be eliminated by dominance editing, we graph the in-basket probability of this offer separately. The remaining offers are then grouped by their ranks in the set of common-standard offers, with higher ranks corresponding with lower final prices. For each of the sets of ranks {2, 3}, {4, 5} and {6, 7, 8}, we graph the average in-basket probability of offers in that set.

Since the defining characteristic of Group 1 is that common-standard offers are inspected first, and since the most time-efficient way to inspect offers is to move them to the basket in batches, it is not surprising that Figure 1a shows high in-basket probabilities for all common-standard offers in the first intervals of clicking time. But from the third interval, an extremely obvious asymmetry emerges. The in-basket probabilities of all common-standard offers except the best decline sharply and then remain close to zero, while that of the best common-standard offer rises to almost 0.9 and then declines gently. In other words, there is a first phase of sorting in which dominance editing is used to eliminate all but one common-standard offer, followed by a second phase in which the surviving common-standard offer is considered alongside individuated-standard offers.

The decline in the in-basket probability of the best common-standard offer during the later intervals of the task implies that significant numbers of subjects put that offer into the basket and then removed it. In some cases, this could have been the result of a correct comparison between that offer and a superior individuated-standard offer. But in fact, it was not uncommon even for objectively optimal common-standard offers to be removed from the basket in PC tasks. Aggregating over all three PC tasks and over all subjects, 30.4 per cent of optimal common-standard offers were moved out of the basket at some stage.¹¹

Figure 1b shows corresponding graphs for individuated-standard offers, again for Group 1 subjects, using the sets of ranks {1}, {2, 3}, {4, ..., 10} and {11, ..., 16}. As one would expect, in-basket probabilities for all individuated-standard offers are very low in the first three intervals of clicking time (during which almost all inspections are of common-standard offers). From then on, individuated-standard offers are added to the basket. From the fifth interval there is systematic

¹¹ For optimal individuated-standard offers, the corresponding proportion was 18.1 per cent. In NC tasks, 21.1 per cent of optimal offers were moved out of the basket. Unsurprisingly, objectively optimal offers were almost never moved out of the basket in AC or benchmark tasks.

sorting in favour of relatively good offers, but this sorting is far less effective than in the case of common-standard offers.

Figures 1c and 1d show the evolution of the in-basket probabilities of common- and individuated-standard offers for Group 2 subjects. Recall that the inspection behaviour of Group 2 subjects did not distinguish between common- and individuated-standard offers. For these subjects, the process of sorting among common-standard offers was necessarily spread over the whole task. It is clear from these graphs that, for both common- and individuated-standard offers, there was systematic sorting in favour of relatively good offers, but that low-ranked offers were less likely to remain in the basket if they were common-standard than if they were individuated-standard.¹²

The most parsimonious explanation of these findings seems to be that both groups of subjects used dominance editing in the sorting process and failed to compensate sufficiently for the resulting bias when comparing surviving common-standard offers with individuated-standard ones. As a result, there was a strong overall bias against common-standard offers. It seems clear that Group 1 subjects used dominance editing to find the best common-standard offer before considering other offers. Our evidence is consistent with the hypothesis that most Group 2 subjects used *continuous* dominance editing. That is, they did not differentiate between common- and individuated-standard offers at the inspection stage, but progressively eliminated dominated offers during the sorting stage, as relations of dominance became apparent.

If this is an accurate characterization of the sorting process, no sharp predictions can be made about whether the *best* shortlisted offer is more likely to reach the final basket if it is common-standard or individuated-standard. The answer to that question depends on a balance of conflicting effects. Because many subjects inspect common-standard offers first, the best shortlisted offer will tend to enter the shopping basket earlier if it is common-standard, and so will be more exposed to the hazard of being eliminated in an inaccurate comparison with an inferior offer. But two factors work in the opposite direction. First, if the best shortlisted offer is common-standard, it can ‘defeat’ seven of the other twenty-three offers in ranking-based comparisons; if it is individuated-standard, it does not have this advantage. Second, some subjects may have some sense of the rationale for combining dominance editing with discrimination in favour of common-standard offers; this rationale might be more salient when the relationship between original and final prices is more complex. So it is perhaps not surprising that, in testing whether being common-standard was an advantage or disadvantage for the best shortlisted offer, we found different answers for different offer types. Since we have relevant data for only three offer types, it would be over-interpretation to try to identify general patterns in these particular results.

¹² Statistical support for this claim is provided by the evidence, referred to above and reported in the Appendix, that sorting by Group 2 subjects discriminated against common-standard offers.

8. Discussion

The behaviour we have observed in this experiment does not support the hypothesis that common pricing standards in retail markets are self-reinforcing. When some but not all offers were priced in a common standard, our subjects were much better able to make price comparisons *within the set of* common-standard offers than *within the set of* individuated-standard offers. If this pattern of behaviour were reproduced in a market, suppliers who priced in common standards would face more price-sensitive consumers than those who used individuated standards. This corresponds with one feature of the Gaudeul-Sugden model described in Sections 2 and 3. Contrary to that model, however, our subjects did not use decision rules that discriminated in favour of common-standard offers. To the contrary, offers were *less* likely to be chosen, other things being equal, if they were priced in common standards. In a market setting, such behaviour would tend to penalize firms which used common standards.

Our experimental design deliberately used mathematically complex price structures. It is well known that people often find it hard to find the lowest price when price structures are complex (Wilson and Waddams Price, 2010; Office of Fair Trading, 2011). It is therefore not surprising that our subjects often failed to find the lowest price in tasks in which there were no common standards (NC tasks) or in which only some offers were priced in common standards (PC tasks). The evidence suggests that this was not due to any lack of effort by subjects or to any weakness of incentives. Recall that in each of these tasks, the median subject chose to spend around five to six minutes inspecting and searching offers, and inspected all or almost all the 24 available offers. Recall too that subjects with more experience of mathematics or economics were more successful in finding low prices. The most natural interpretation of these findings is that subjects devoted considerable effort to trying to compare final prices, but these efforts were sometimes defeated by the difficulty of the mental arithmetic involved.

What is more interesting is that our subjects gained little or no benefit from the presence of common standards in PC tasks, despite the opportunities this offered for simplifying the problem of searching for low prices. There is a sense in which subjects expended *too much* effort in these tasks: they failed to use a simple decision rule (the ‘largest common standard’ or LCS heuristic) that was clearly within their cognitive capacities and which would have allowed them to find lower prices with considerably less expenditure of time and effort. Indeed, the majority of subjects (those that we have classified as ‘Group 1’) used a decision rule whose first stage replicated the LCS heuristic by finding the best common-standard offer, but then wasted both time and money by comparing this offer with offers priced in individuated standards.

Our subjects may have been reluctant to use (or failed to consider using) the LCS heuristic because of the counter-intuitiveness of a strategy that looks at only one in every three of the available

offers. Most of our subjects would have known how to carry out the arithmetic operations which, if applied correctly to all 24 offers, would be certain to identify the lowest price. One possible interpretation of our findings is that subjects were overconfident in their ability to avoid arithmetic errors. There is a long-standing psychological literature on the reliability of people's judgements of confidence in their own knowledge, beliefs, skills and memory. This research has found a prevailing but not universal bias towards overconfidence, particularly in relation to complex tasks (e.g. Adams and Adams, 1961; Oskamp, 1965; Lichtenstein and Fischhoff, 1977). Economists and management scientists have been interested in overconfidence as a cause of excess market entry (Camerer and Lovo, 1999), excess financial trading (Barber and Odean, 2001), and excess investment by CEOs in their firms' own projects (Malmendier and Tate, 2005). It is possible that overconfidence by consumers about their own cognitive abilities may make them more vulnerable to pricing practices which make rival offers difficult to compare.

An alternative (or complementary) explanation of our findings is that the operation of eliminating transparently dominated options has an obvious rationale and is cognitively undemanding. It requires much more sophisticated thinking to recognize the information contained in the fact that a surviving option dominates other offers that are no longer under consideration. Viewed in this perspective, it is perhaps not so surprising that our subjects used dominance editing in the process of constructing a final 'shopping basket' or shortlist, but then did not compensate sufficiently for the resulting bias in the composition of the basket.

It is possible that individuals would be more willing to ignore individuated-standard offers in situations in which they could rationalize this practice without thereby acknowledging their own cognitive limitations. This might be the case, for example, if a decision had to be made within an externally-imposed time constraint, or if there was an explicit monetary cost for inspecting each offer (something that is relatively rare in real-world shopping). The nature of the mental operations required to identify the best offer may also be significant. For example, individuals might be more relaxed about acknowledging deficiencies in spatial judgement than in arithmetic ability (or vice versa). These conjectures may help to explain differences between our results and those of an experiment reported by Crosetto and Gaudeul (2012), which found some evidence of the use of the LCS heuristic. In a typical task in Crosetto and Gaudeul's experiment, a subject had to make a decision problem within a fixed time constraint. The problem was to purchase 'paint' to cover a fixed square area. The prices of alternative paint products were described in terms of the cost of covering areas of various shapes and sizes, presented visually; in some choice problems, two or more products were priced in terms of areas with the same shape and size, thus creating a common standard.

Nevertheless, our conclusion has to be that we have not found support for the hypothesis that consumers have a basic propensity to construct shortlists that discriminate in favour of common-standard offers. Our findings draw attention to three factors – the psychological salience of editing operations which eliminate dominated options, the cognitive sophistication needed to recognize and correct the biases induced by these operations, and the possibility that individuals are overconfident in their ability to make complex price comparisons – which work against this form of shortlisting.

References

- Adams, Joe K., and Pauline A. Adams (1961). Realism of confidence judgments. *Psychological Review* 68: 33–45.
- Barber, Brad M. and Terrance Odean (2001). Boys will be boys: gender, overconfidence, and common stock investment. *Quarterly Journal of Economics* 116: 261–292.
- Brandstätter, Eduard, Gerd Gigerenzer and Ralph Hertwig (2006). The priority heuristic: making choices without trade-offs. *Psychological Review* 113: 409–432.
- Camerer, Colin and Dan Lovallo (1999). Overconfidence and excess entry: an experimental approach. *American Economic Review* 89: 306–318.
- Crosetto, Paulo and Alexia Gaudeul (2012). Do consumers prefer offers that are easy to compare? An experimental investigation. *Jena Economic Research Papers* #2011-044.
http://pubdb.wiwi.uni-jena.de/pdf/wp_2011_044.pdf
- Diamond, Peter A. (1971). A model of price adjustment. *Journal of Economic Theory*, 3(2), 156–168.
- Eliasz, Kfir, Michael Richter and Ariel Rubinstein (2011). Choosing the two finalists. *Economic Theory*, 46(2): 211–219.
- Eliasz, Kfir and Ran Spiegler (2011). Consideration sets and competitive marketing. *Review of Economic Studies* 78(1): 235–262.
- Ellison, Glenn (2005). A model of add-on pricing. *Quarterly Journal of Economics* 120: 585-637.
- Ellison, Glenn and Sara Fisher Ellison (2009). Search, obfuscation, and price elasticities on the internet. *Econometrica* 77: 427–52.
- Fischbacher, Urs. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 102: 171–178.
- Gabaix, Xavier and David Laibson (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quarterly Journal of Economics* 121: 505-540.

- Gaudeul, Alexia and Robert Sugden (2012). Spurious complexity and common standards in markets for consumer goods. *Economica* 79: 209–225.
- Goldstein, Daniel G. and Gerd Gigerenzer (2002). Models of ecological rationality: the recognition heuristic. *Psychological Review* 109: 75–90.
- Hauser, John R. (2010). Consideration-set heuristics. Forthcoming in *Journal of Business Research*.
- Hauser, John R. and Birger Wernerfelt (1990). An evaluation cost model of consideration sets. *Journal of Consumer Research* 16: 393-408.
- Huber, Joel, John W. Payne and Christopher Puto (1982). Adding asymmetrically dominated alternatives: violations of regularity and the similarity hypothesis. *Journal of Consumer Research* 9: 90–98.
- Kahneman, Daniel and Amos Tversky (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47: 263–292.
- Lichtenstein, Sarah and Baruch Fischhoff (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance* 20: 159–183.
- Malmendier, Ulrike and Geoffrey Tate (2005). CEO overconfidence and corporate investment. *Journal of Finance* 60: 2661–2700.
- Manzini, Paola and Marco Mariotti (2007). Sequentially rationalizable choice. *American Economic Review* 97: 1824–1839.
- Menger, Carl (1892). On the origins of money. *Economic Journal* 2: 239–255.
- Office of Fair Trading (2011). Consumer behavioural biases in competition: a survey. Report by London Economics in association with Steffen Huck and Jidong Zhou.
- Oskamp, Stuart (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology* 29: 261–265.
- Payne, John W. (1976). Task complexity and contingent processing in decision making: an information search and protocol analysis. *Organizational Behavior and Human Performance* 16: 366–387.
- Rubinstein, Ariel (1988). Similarity and decision-making under risk (is there a utility theory resolution to the Allais Paradox? *Journal of Economic Theory* 46: 145–153.
- Salop, Steven and Joseph Stiglitz (1977). Bargains and ripoffs: A model of monopolistically competitive price dispersion. *Review of Economic Studies*, 44: 493–510.

- Schotter, Andrew (1981). *The Economic Theory of Social Institutions*. Cambridge: Cambridge University Press.
- Shafir, Eldar, Itamar Simonson, and Amos Tversky (1993). Reason-based choice. *Cognition* 49: 11–36.
- Simon, Herbert (1956). Rational choice and the structure of the environment. *Psychological Review* 63: 129–138.
- Spiegler, Ran (2006). Competition over agents with boundedly rational expectations. *Theoretical Economics* 1: 207–31.
- Sugden, Robert (2004). *The Economics of Rights, Co-operation and Welfare*. Basingstoke: Palgrave. First edition 1986.
- Tversky, Amos (1969). Intransitivity of preferences. *Psychological Review* 76: 31–48.
- Tversky, Amos (1972). Elimination by aspects: a theory of choice. *Psychological Review* 79: 281–299.
- Wilson, Chris M. and Waddams Price, Catherine (2010). Do consumers switch to the best supplier? *Oxford Economic Papers* 62: 647–668.

Table 1
Properties of the tasks

Task	Task type	Offer type	Number of common standard offers	Number of individuated standard offers
1	Benchmark	P	24	0
2	AC	P*D	24	0
3	AC	P*D+A	24	0
4	AC	P*D1*D2	24	0
5	NC	P*D	0	24
6	NC	P*D+A	0	24
7	NC	P*D1*D2	0	24
8	PC	P*D	8	16
9	PC	P*D+A	8	16
10	PC	P*D1*D2	8	16

Table 2

Time used, rank of chosen offer and frequency of optimal choice in each task

Task	Task type	Offer type	Time used (seconds)		Ranked payment		Proportion of optimal choices (%)
			Mean (SD)	Median	Mean (SD)	Median	
1	Benchmark	P	158 (104)	125	1.88 (3.62)	1	89.5
2	AC	P*D	177 (145)	133	2.26 (4.37)	1	86.0
3	AC	P*D+A	187 (132)	143	2.33 (4.57)	1	84.8
4	AC	P*D1*D2	183 (121)	138	2.82 (5.50)	1	81.9
5	NC	P*D	344 (226)	305	5.32 (5.34)	3	26.3
6	NC	P*D+A	363 (225)	328	5.33 (5.52)	3	27.5
7	NC	P*D1*D2	398 (296)	355	7.58 (6.73)	5	20.5
8	PC	P*D	342 (256)	289	4.51 (5.18)	2	36.8
9	PC	P*D+A	354 (245)	304	5.51 (5.44)	3	21.6
10	PC	P*D1*D2	376 (243)	363	7.13 (6.51)	5	19.3

There are 171 observations for each task. 'Ranked payment' is a linear transformation of the final price of the chosen offer. It takes the value 1 at the lowest final price (£20.27) and 24 at the highest final price (£28.90). SD = standard deviation.

Table 3

Numbers of common and individuated-standard offers inspected in each task

Task	Task type	Offer type	Number of CS offers			Number of IS offers			<i>p</i>
			inspected			inspected			
			Mean (SD)	Median	%	Mean (SD)	Median	%	
1	Bench- mark	P	22.6 (4.52)	24	94.3				
2	AC	P*D	22.3 (4.88)	24	93.0				
3	AC	P*D+A	22.3 (4.83)	24	92.7				
4	AC	P*D1*D2	22.4 (4.53)	24	93.1				
5	NC	P*D				19.9 (6.33)	24	83.0	
6	NC	P*D+A				19.8 (6.28)	24	82.4	
7	NC	P*D1*D2				18.4 (6.61)	22	76.6	
8	PC	P*D	7.3 (1.78)	8	91.2	13.5 (4.02)	16	84.1	<0.001***
9	PC	P*D+A	7.1 (2.12)	8	88.3	12.9 (4.46)	16	80.9	<0.001***
10	PC	P*D1*D2	7.0 (2.12)	8	87.5	12.3 (4.61)	15	76.8	<0.001***

There are 171 observations for each task. CS = common standard; IS = individuated-standard; SD = standard deviation. The final column reports the *p*-value for a signed rank test of the null hypothesis that, in PC tasks, the proportion of inspected offers is the same for CS and IS. *, **, and *** (respectively: #, ##, and ###) denote differences in the direction predicted (respectively: not predicted) by Hypothesis 2 that are significant at the 10, 5 and 1 per cent levels.

Table 4

Frequency of choice of inspected common-standard offers in PC tasks

Task	Offer type	Choices of common standard offers	%	CS offers inspected as % of all offers inspected	<i>p</i>
8	P*D	50	29.2	35.1	0.128
9	P*D+A	29	17.0	35.5	<0.001###
10	P*D1*D2	46	26.9	36.3	0.011##

There are 171 observations for each task. CS = common-standard; IS = individuated-standard. The final column reports the *p*-value for a two-tail binomial test of the null hypothesis that the probability that a CS offer is chosen is equal to the observed ratio of inspected CS offers to total inspections (aggregated over all subjects). *, **, and *** (respectively: #, ##, and ###) denote differences in the direction predicted (respectively: not predicted) by Hypothesis 3 that are significant at the 10, 5 and 1 per cent levels.

Table 5

Frequency of choice of best inspected offer in PC tasks

Task	Offer type	If BI offer is common standard			If BI offer is individuated standard			<i>p</i>
		<i>n</i>	Number of subjects choosing BI offer	%	<i>n</i>	Number of subjects choosing BI offer	%	
8	P*D	61	27	44.3	110	42	38.2	0.438
9	P*D+A	65	13	20.0	106	36	34.0	0.050#
10	P*D1*D2	61	26	42.6	110	20	18.2	<0.001***

There are 171 observations for each task. BI offer = best inspected offer. The final column reports the *p*-value for a *z*-test of the null hypothesis that the probability that the BI offer is chosen is independent of whether that offer is common- or individuated standard. *, **, and *** (respectively: #, ##, and ###) denote differences in the direction predicted (respectively: not predicted) by Hypothesis 4 that are significant at the 10, 5 and 1 per cent levels.

Table 6**Offers in final basket as proportion of offers inspected: PC tasks**

Task	Offer type	Number of CS offers inspected	Number of IS offers inspected	Number of CS offers in final basket	Number of IS offers in final basket	Proportion of inspected CS offers in final basket (%)	Proportion of inspected IS offers in final basket (%)	<i>p</i>
8	P*D	7.30 (1.78)	13.46 (4.03)	0.79 (0.99)	3.23 (2.48)	15.3 (24.84)	29.2 (27.75)	<0.001###
9	P*D+A	7.06 (2.12)	12.94 (4.45)	0.79 (1.40)	2.87 (2.18)	15.4 (27.65)	27.6 (26.99)	<0.001###
10	P*D1*D2	7.00 (2.12)	12.29 (4.62)	0.94 (1.30)	3.70 (2.59)	17.3 (26.41)	36.4 (29.64)	<0.001###

Entries in table are means (standard deviations in parentheses). CS = common standard; IS = individuated standard. In task 8 (respectively: 9, 10), 0 (respectively: 3, 3) subjects inspected no CS offers, and 1 (respectively: 2, 1) subjects inspected no IS offers. These subjects are excluded from columns 7 and 8 and from the statistical tests. The final column reports the *p*-value for a signed rank test of the null hypothesis that the proportions of inspected CS and IS offers that ended up in the final basket are the same. *, **, and *** (respectively: #, ##, and ###) denote differences, significant at the 10, 5 and 1 per cent levels, in the direction consistent (respectively: not consistent) with the hypothesis that this proportion is higher for CS offers.

Table 7

Frequency with which best inspected offer is in final basket: PC tasks

Task	Offer type	If BI offer is common-standard			If BI offer is individuated-standard			<i>p</i>
		<i>n</i>	Number of subjects with BI offer in the final basket	%	<i>n</i>	Number of subjects with BI offer in the final basket	%	
8	P*D	61	40	65.6	110	74	67.3	0.821
9	P*D+A	65	25	38.5	106	61	57.6	0.015##
10	P*D1*D2	61	47	77.1	110	50	45.5	<0.001***

There are 171 observations for each task. BI offer = best inspected offer. The final column reports the *p*-value for a *z*-test of the null hypothesis that the probability that the BI offer is in the final basket is independent of whether that offer is common-standard or individuated-standard. *, **, and *** (respectively: #, ##, and ###) denote differences, significant at the 10, 5 and 1 per cent levels, in the direction consistent (respectively: not consistent) with the hypothesis that the BI offer is more likely to be in the final basket if it is common-standard than if it is individuated-standard.

Table 8

Frequency with which best offer in final basket is chosen: PC tasks

Task	Offer type	If BFB offer is common-standard			If BFB offer is individuated-standard			<i>p</i>
		<i>n</i>	Number of subjects choosing BFB offer	%	<i>n</i>	Number of subjects choosing BFB offer	%	
8	P*D	58	41	70.7	113	72	63.7	0.362
9	P*D+A	38	22	57.9	133	87	65.4	0.395
10	P*D1*D2	64	35	54.7	107	55	51.4	0.677

There are 171 observations for each task. BFB offer = best offer in final basket. The final column reports the *p*-value for a *z*-test of the null hypothesis that the probability that the BFB offer is chosen is independent of whether BFB is common-standard or individuated-standard. *, **, and *** (respectively: #, ##, and ###) denote differences, significant at the 10, 5 and 1 per cent levels, in the direction consistent (respectively: not consistent) with the hypothesis that the BFB offer is more likely to be chosen if it is common-standard than if it is individuated standard.

Table 9

Distribution of first eight inspections in PC tasks

Task	Offer type	Number of subjects who inspected m common-standard offers: $m =$								
		0	1	2	3	4	5	6	7	8
8	P*D	4	17	16	17	6	4	8	18	81
9	P*D+A	4	16	13	17	11	3	5	20	82
10	P*D1*D2	4	19	24	10	15	3	7	13	76

There are 171 observations for each task.

Figure 1

Evolution of contents of shopping basket for task 8

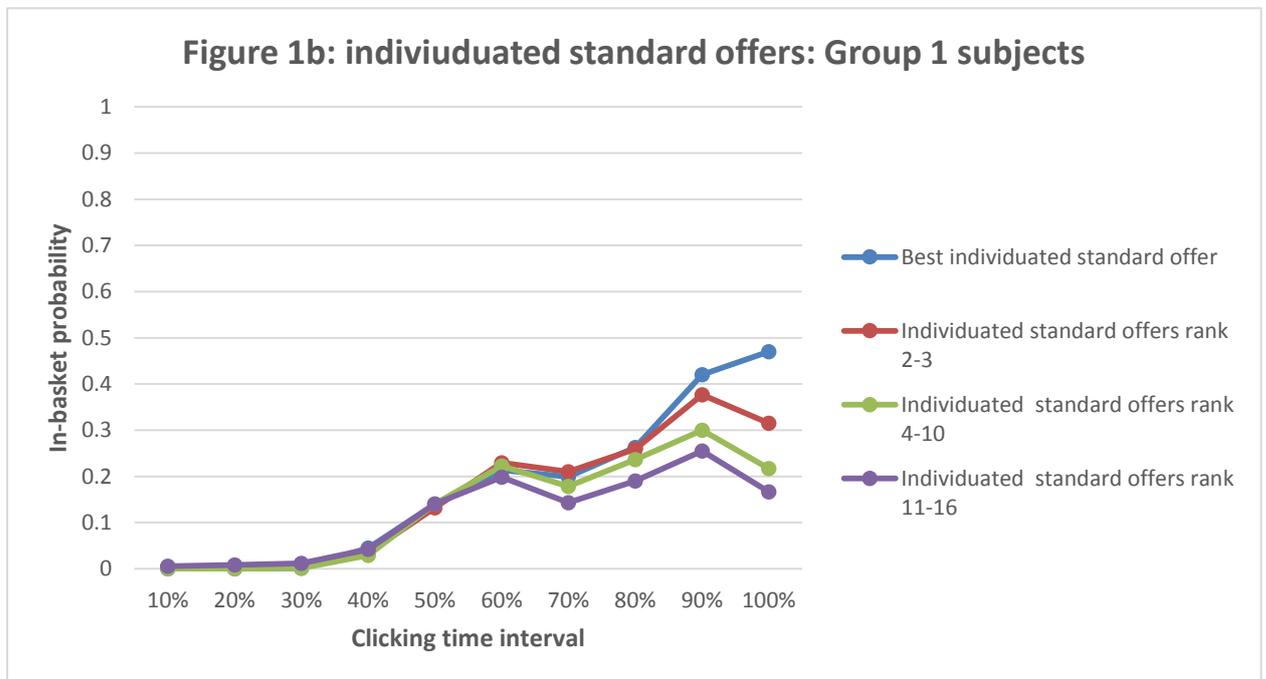
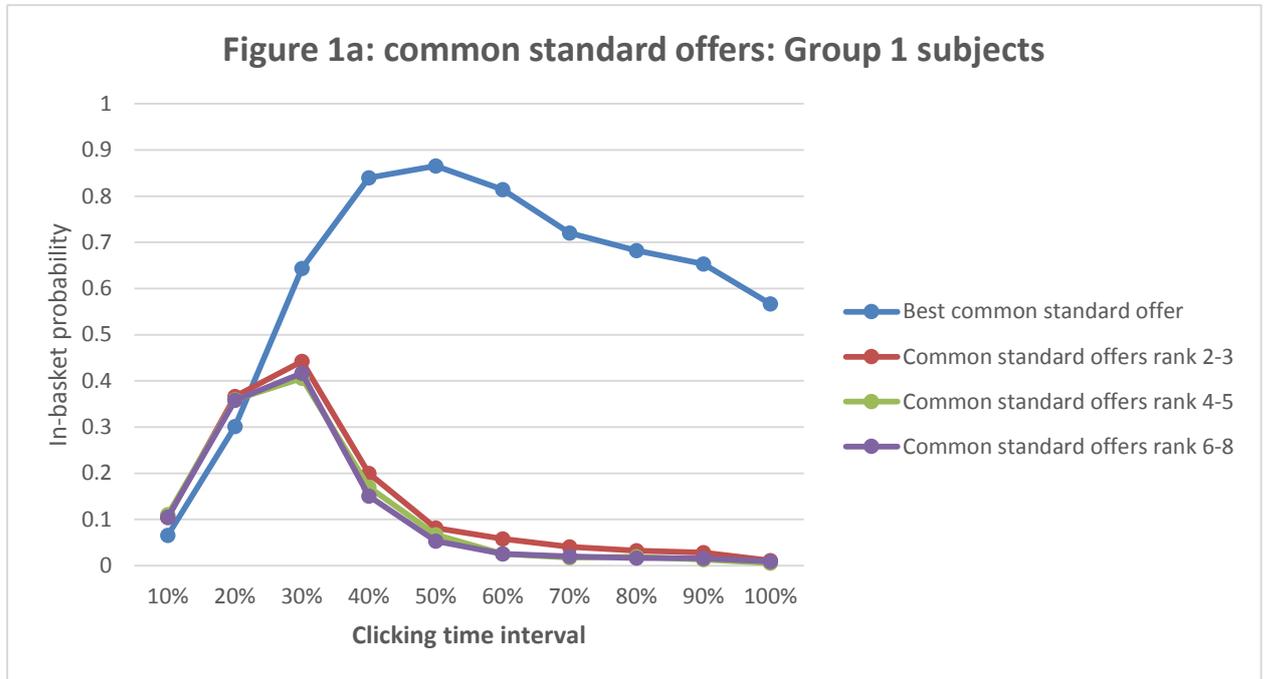


Figure 1c: common standard offers: Group 2 subjects

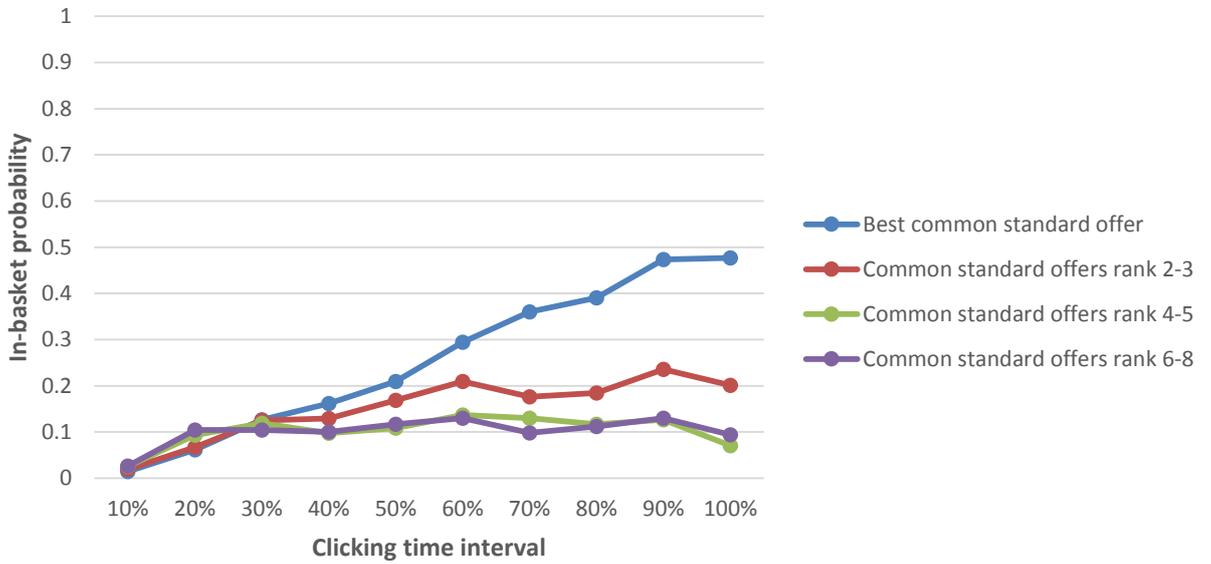


Figure 1d: individuated standard offers: Group 2 subjects

