Journal of Banking and Finance 121 (2020) 105955

Contents lists available at ScienceDirect

Journal of Banking and Finance

journal homepage: www.elsevier.com/locate/jbf

Flicking the switch: Simplifying disclosure to improve retirement plan choices

S. Thorp[a,*], H. Bateman[b], L.I. Dobrescu[c], B.R. Newell[d], A. Ortmann[e]

[a] Discipline of Finance, The University of Sydney Business School, University of Sydney, NSW 2006, Australia
[b] CEPAR and School of Risk and Actuarial Studies, UNSW Sydney, NSW 2052, Australia
[c] CEPAR and School of Economics, UNSW Sydney, NSW 2052, Australia
[d] School of Psychology, UNSW Sydney, NSW 2052, Australia
[e] School of Economics, UNSW Sydney, NSW 2052, Australia

ARTICLE INFO

Article history:
Received 20 August 2019
Accepted 4 September 2020
Available online 19 September 2020

JEL codes:
G51
G28
G41

Keywords:
Information disclosure
Pensions
Choice experiment

ABSTRACT

Standardized information disclosures aim to help people compare complex financial products and make better choices. We investigate the extent to which information shown in a regulator-mandated dashboard helps retirement savers choose between alternative pension plans. We conduct incentivized experiments that collect participants' repeated choices between two pension plans using the mandatory dashboard, and subsequently test whether an even simpler dashboard improves choices, and by how much. Participants switch quickly from a high- to a low-fee pension plan when they see explicit nominal fees but are significantly more confused by percentage fees and adjust slower. When differences between plan performance arise from gross returns, not fees, we find that complex information formats can seriously hinder participants' recognition and reactions. We present a Bayesian updating model which estimates the relative noisiness of the signals from fees and gross returns across different treatments and use this model to show how better information presentation raises retirement savings.

1. Introduction

Many consumers make costly mistakes because they find financial products hard to understand and difficult to compare. Market forces alone will not always correct these inefficiencies wherever financial corporations can benefit from poor financial decisions (Campbell et al., 2011). Regulators promote standardized product disclosures to support consumers' financial decisions (Loewenstein et al., 2014), however even these can be used in unexpected ways, with detrimental results (Choi et al., 2010; Navarro-Martinez et al., 2011; Agarwal et al., 2014; Jones et al., 2015; Bateman et al., 2016).

Here we investigate how pension plan participants respond to performance information presented in a dashboard implemented by the Australian financial regulator in 2014. This a particularly informative and timely setting for several reasons. First, the complexity of retirement savings decisions and consequent confusion among participants (Benartzi and Thaler 2001; Brown et al., 2007; Bhandari and Deaves 2008; Chetty et al., 2014) weakens competitive pressures on pension plan providers. As a result, some retirement plans retain under-performing, affiliated funds (Pool et al., 2016) and offer high-fee or dominated funds on investment menus (Ayres and Curtis, 2015). Second, this lack of competitive pressure is an acute problem for plan participants in Australia. Retirement contributions of 9.5% of earnings are compulsory for most Australian workers and Australian plans (called "superannuation funds") now hold the world's second largest pool of defined contribution (DC) savings after the U.S. (Vanguard, 2017). Sectoral scale, however, has not guaranteed efficiency. A major public review described competition between Australian retirement plans as "superficial", remarking that the 8% of account holders that ended up in under-performing plans were likely to accumulate 50% less by retirement than they would in stronger-performing plans (Productivity Commission 2018, p. 2). Third, we conducted our study in 2014–15, just as the regulator required pension plans to adopt - and prominently place on their website - a prescribed performance "dashboard" for their DC strategies. We thus conduct an opportune test of a "real-world" intervention, designed to benefit millions of DC participants in Australia, and with the potential to inform pension participants, providers and regulators in many other countries.

We investigate the effectiveness of the plan dashboard through a program of incentivized online experiments. Specifically, we offer

* Corresponding author.
E-mail addresses: susan.thorp@sydney.edu.au (S. Thorp), h.bateman@unsw.edu.au (H. Bateman), dobrescu@unsw.edu.au (L.I. Dobrescu), ben.newell@unsw.edu.au (B.R. Newell), a.ortmann@unsw.edu.au (A. Ortmann).

https://doi.org/10.1016/j.jbankfin.2020.105955
0378-4266/

Australian retirement savers a sequence of 20 choices between two hypothetical DC plans that differ in systematic ways but are otherwise identical. Each plan's characteristics and performance are shown in one of two types of dashboard, either a "standard" one-page summary that exactly recreates the format and content stipulated by the Australian regulator, or a further "simplified" version that presents essentially the same content in a different format. While the dashboards we test include information on fees, net returns and investment risk, we focus only on differences between plans due to fees or (implied) gross returns because these are basic signals of plan quality. These three design features, that is, a sequence of 20 annual plan choices, standard versus simplified dashboard format, and fee-sourced versus gross-return-sourced performance differences, let us: (i) see what information from the dashboard most influences participants; (ii) estimate how much uncertainty participants associate with fee information as compared with gross returns information, and with the standard dashboard as compared with the simplified dashboard, and (iii) evaluate the impact of information sources and dashboard format on participants' final savings.

In the experiments, participants choose between the two plans in each of 20 "years" that represent the repeated choices over plans that people typically make while accumulating retirement savings during their working life. At the start, one plan's net-of-fee returns clearly dominates the other, but as participants proceed through the 20 years of choices, we gradually update the relative performance of the plans. This is done by changing either fees or gross returns holding expected investment risk constant, in an attempt to induce participants to switch towards the alternative plan. Australian retirement savers must be able to recognize and act on these performance signals if the dashboard is to facilitate competitive discipline on retirement plans, as the regulator intends (Australian Securities and Investments Commission (ASIC), 2013, Australian Securities and Investments Commission (ASIC), 2014). By incorporating a sequence of updating performance signals, we can test whether the dashboard can help reduce inertia among participants who stick with under-performing plans (Productivity Commission, 2018).

Since we set expected investment risk to be constant and the same for both plans, experiment participants should select the plan that offers the highest expected net return. Participants can combine fees and net returns information to infer that the difference in realized net returns between plans is due either to differences in fees or to differences in realized gross returns. We design the experiment to ensure that differences in realized net returns are due to one or the other, but not both simultaneously, and we build in a crossing point around the middle of the sequence of 20 choices where the initially out-performing plan is equalled, and then subsequently increasingly out-performed by the alternative plan. Participants in treatments where performance differences are caused by fee differences should choose the lowest-fee (highest net return) plan each year. Participants in treatments where performance differences show up in gross returns, not fees, may treat gross returns as a noisy signal, update expected returns cautiously, and take longer to switch plans. We test whether participants can infer these differences from the dashboard over repeated rounds, allowing for the possibility that they might update expected net returns differently when fees change than when realized gross returns change. We measure differences between information formats that condition on this possibility so that our findings about the clarity of the dashboard do not depend on us knowing (or setting) participants' prior beliefs about expected net returns.

Despite the overwhelming evidence that investors should pay attention to fees (Gil-Bazo and Ruiz-Verdú, 2009), information disclosure studies find that fee information is often overlooked or misunderstood (Barber et al., 2005; Choi et al., 2010; Beshears et al., 2011; Fisch and Wilkinson-Ryan 2014) especially by less able investors (Grinblatt et al., 2015). Investor errors may be exacerbated by particular structures of mutual fund fees (Anufriev et al., 2019). Our online experiments test (i) whether participants respond to fee information and switch to an otherwise equivalent lower fee plan when gross returns and risk for both plans are equal, and (ii) which of two equivalent fee formats (nominal vs. annual percentages for a $50,000 balance) are better understood. We find that most participants understand and react to "all-in" fee information (when shown in dollar amounts), promptly choosing the cheaper plan. The same fee information presented as (scale compatible) annual percentages of assets, however, elicits a slower and more tentative reaction.

For gross returns, we present three sets of results. First, we expect participants to update their plan performance expectations at every round as new gross returns information arrives, always choosing the plan with higher expected net returns. Those participants who treat gross returns as a noisy signal of plan skill may delay switching until they have noticed several "years" of outperformance, but our experimental design, where differences between plans change monotonically, ensures that they should switch between plans only once, independent of their prior beliefs about gross returns signals. We find that simplifying the dashboard indeed delivers significantly higher rates of one-time switches.

Second, we employ a Bayesian model to estimate and compare the noise each participant attaches to identical fees and gross returns when presented in the regulated and the simplified dashboard, respectively. We can compare noise estimates for different information formats allowing for different prior beliefs. Simplification reduces the estimated noise in gross returns information by more than 40%. We also manipulate – still within the same dashboard context – the relative gross return volatility and compare the estimated noise associated with gross returns at low and high volatilities. This sheds further light on whether people's ability to track gross returns depends on the underlying volatility, a general uncertainty about the predictive value of gross returns, and/or on complexity-induced confusion.

Third, we test whether participants react differently to relative performance information due to rising, versus falling, fees or gross returns. To do so, we vary performance information so that in one condition the dominant plan's performance gradually deteriorates (relative to the benchmark plan), and in the other an underperforming plan gradually improves. We find that performance differences are harder to detect and respond to if they are related to improving (rather than worsening) relative performance. This experimental result aligns with empirical findings that mutual fund investors opt out of underperforming funds less readily than they opt into highly performing ones (Sirri and Tufano, 1998). We further show, however, that this tendency is made worse by complex information.

Unlike many studies that focus on the effect of different risk formats (e.g., Kaufmann et al., 2013), we keep risk information constant across plans. We do so, first, because most Australian pension plans offer a 70:30 growth:defensive asset mix and very similar long-run risk and return targets for their default asset allocation (to which the dashboard test applies). Hence, participants who are contemplating a switch from their current plan are likely to compare options with the same risk target. Second, we can experiment directly on the effect of fee and gross returns information while controlling for risk. By letting performance differences in each treatment be driven by *either* fees *or* gross returns, we can thus disentangle the effect of simplifying the information on each factor for plan choice. Our setting mimics real outcomes for retirement saving plans in Australia where plans with the same general risk profile produce persistent performance differences (Productivity Commission 2018).

Overall, we show that the simplified dashboard delivers gross returns information more effectively than the standard dashboard, leading to significantly higher retirement account balances. However, fee information is better understood in the standard setting, when shown as an "all-in" dollar amount and where returns are less salient.

Our results add to the literature on the effects on financial decisions of information disclosure and complexity, while also proposing and evaluating specific remedies. Theoretical studies suggest that financial firms use information complexity to shroud product attributes and confuse retail investors (Gabaix and Laibson, 2006; Carlin, 2009; Carlin and Manso, 2010). These predictions have been confirmed empirically using administrative data (Henderson and Pearson, 2011; Célérier and Vallée, 2017), with experimental evidence from trading further showing that complexity creates adverse selection and extra uncertainty (Carlin et al., 2013; Grubb, 2015; Kalayci, 2015, 2016). Here we present new experimental evidence of plan participants' reactions to prescribed information about key plan features and subsequent simplifications, rather than evaluating the strategic behavior of financial firms or traders. We also go further by assessing and quantifying the beneficial or detrimental effect of information simplifications on savings outcomes.

We also contribute to the literature on regulated product disclosures by analyzing the joint impact of key plan features contained in a prescribed disclosure format. Until recently, such elements have rarely been considered jointly (see Diacon and Hasseldine, 2007; Gillis, 2015; Colaert, 2016). Indeed, most studies have looked at presentation formats of risk, fees, and returns separately, sometimes comparing summary against comprehensive disclosures (e.g., Kozup et al., 2008; Beshears et al., 2011; Walther, 2015; Anufriev et al., 2019).

In somewhat related work, Anufriev et al. (2019) (see also Anufriev et al., 2016; 2018, 2019) explore systematically a variant of heterogeneous-agents learning-to-forecast models under various conditions (e.g., negative feedback vs. positive feedback). Our point of departure differs: we hypothesize the real-world template to be too complicated and we hence investigate simplifications. In its essence, our paper is about framing effects in an important real-life setting and our model is custom-made to tease out the relative noisiness of the fee and gross return signals from across the different treatments.

2. Institutional setting

In Australia, regular flows of mandatory savings (at a minimum contribution rate of 9.5% of earnings) have built the pension sector to more than US$2 trillion (APRA, 2018) - the world's second largest pool of DC savings after the U.S. (Vanguard, 2017). While rated as one of the best pension systems in the world (Mercer 2017), compulsion and inertia combine to weaken the competitive pressure on plans and foster inefficiencies (Productivity Commission 2018).[1] For instance, annual administration and investment fees for very similar DC plans vary by more than 100%, and investment offerings with the same risk and return goals report a large range of investment returns (APRA, 2017). These discrepancies are partly due to very few people opting out of poorly performing plans: less than a third of members ever opt out of the default plan selected by their employer (Chant et al., 2014; Minifie et al., 2015), fewer than 10% of members switch plan providers in a typical year despite being free to do so (Productivity Commission, 2016) and over 40% of plan members have more than one account and thus pay redundant charges. High and low performing plans needed to be clearly identifiable to participants to motivate switches between products based on simple comparisons (Productivity Commission, 2018).

Comparing alternative pension plans is particularly challenging and one of the reasons why participants do not opt out of under-performing plans or consolidate all savings into one plan (Productivity Commission, 2016). To address this impediment, from 2014 onwards, the Australian regulator required pension plans to adopt - and prominently place on their website - a prescribed one-page disclosure format known as a "dashboard" for its new regulated default, called "MySuper". (MySuper products are (purportedly) low-cost DC savings vehicles that conform to investment strategy, service provision, and fee regulations.) This standard "dashboard" was designed by regulators in consultation with industry to assist potential plan participants to compare (MySuper) default products easily by their returns, risks, and fees (Cooper, 2010; Commonwealth of Australia, 2013; APRA 2015;). In practice, default product providers must prominently place on their website[2] an up-to-date dashboard displaying:

- *Return target* - calculated as the mean annualized estimate of the percentage rate of (net) return above the CPI's growth over the next ten years.
- *Returns* - calculated as the return for each of the past 10 financial years net of administration and advice fees, costs, and taxes from the net investment return.
- *Comparison between return target and returns* - shown on a graph that includes (i) the net returns of the last 10 financial years (shown in columns as percentage rate of return), and (ii) the moving average return target and moving average net return (both shown as lines).
- *Level of investment risk* - presented via a standard risk-measure format, with investment risk shown as the expected number of negative returns for the product over 20 years and accompanied by a scale ranging from very low to very high (FSC and ASFA, 2011).
- *Fees and other costs* - calculated as the dollar amount of fees and other costs for an account balance of A$50,000 (Commonwealth of Australia, 2013; ASIC, 2014; APRA, 2015).

The regulator recommended that the standard dashboard include two "warnings" (see Supplementary materials A), namely: (i) the *return target* box contains the statement "Future returns cannot be guaranteed. This is a prediction.", and (ii) the returns graph warns that "[*p*]ast performance is not necessarily an indicator of future returns". Interestingly, this dashboard also shows the 10-year average return net of fees without any explicit warning, and reports fees without any hint they could change in the future.

We use the standard dashboard format as a template for our tests and as a benchmark for our information simplifications. While this prescribed disclosure is – as far as we know – the first of its kind for DC pension plans, this format is also consistent with the types of reforms called for in 401(k) reporting in the U.S. (Ayres and Curtis, 2015). Moreover, a recent public Inquiry (Productivity Commission 2018) recommended that the Australian regulator host a website where participants can easily compare different plan dashboards and, if desired, to switch, exactly like in our experimental setting (see more below). Such side-by-side comparisons of prescribed disclosures are not completely different from

[1] Note that while workers are automatically enrolled in a superannuation fund, only those who do not choose a plan for themselves are defaulted into the plan chosen by their employer.

[2] In practice, dashboards were initially not easy for participants to find on plan websites, often not prominently displayed and requiring several "clicks" before landing (Australian Securities and Investments Commission ASIC, 2017). While we cannot be sure that experiment participants had not previously seen a dashboard, it would be very unlikely.

what we see in other contexts, for instance when choosing credit cards.[3]

3. Experiment overview

The regulator intended that the plan dashboard should allow participants to make easy and clear comparison between similar plans and that consequently participants would be induced by performance information to switch away from a (default) plan that persistently under-performed. The dashboard designers expected that more, and quicker, switching out of underperforming plans would reduce losses to plan participants and raise competitive pressure in the sector (Productivity Commission 2018). We thus designed the experiments to test whether participants could see differences between plans both initially, and importantly, as they evolved over time, and whether the dashboard would enable switching after a sequence of signals when doing so was very easy. If the dashboard cannot do this, it is unlikely to improve outcomes for plan participants. We then compare the standard dashboard with a simplified version.

Between July 2014 and October 2015, we conducted seven separate treatments (T1-T7) involving over 1800 pension plan participants. We recruited participants from the *Pureprofile* representative online panel of over 600,000 Australians, all of whom were 18+ and enrolled in a pension plan. We sampled 250 to 286 people for each treatment, with a 50:50 split by gender, and we approximated population age proportions for cohorts of 18–34, 35–49, and 50–64 years.

The panel provider recruited participants via email and invited them to click a link to the consent page. We then asked them to read an information page that explained the study purpose, survey structure, confidentiality, compensation, and offered feedback sources. Consent to participate moved participants to the screening questions and if eligible, they continued to the survey. Comparing survey participants with Australian Census data indicates that they are better educated, slightly more likely to work full time and earn higher income than the general population (see Supplementary materials B).

3.1. Incentives

We paid incentives to encourage participants to complete the survey, be as discerning as possible in the task, and as accurate as possible in task comprehension and financial literacy quizzes. The completion payment plus bonus incentive applied to all participants, apart from a hold-out test group of 138 participants in treatment 1, who were compensated only for completing the survey.[4] The consent form specified that participants would be compensated in two ways. First, on completing the survey, they would receive around A$4 worth of *Pureprofile* points, redeemable for cash or goods and services. Second, participants could earn up to an additional A$4 bonus depending on the quality of the answers they gave in one of three randomly selected sections of the survey. The bonus payments were either (i) the average net returns to the plans they selected in the choice task, applied to a A$3 account balance,[5] (ii) the proportion of correct answers to the comprehension questions multiplied by A$3, or (iii) the proportion of correct answers to the questions on financial literacy, numeracy, and pension system knowledge multiplied by A$3. (The hold-out group in treatment 1 did not get the offer of the bonus payment.) The average bonus payment was A$2.18 with a standard deviation of A$1.10. At the end of the survey, we computed the bonus earned by each participant, explained how it was calculated, and showed them the amount. The median participant took less than 20 minutes to complete the survey.[6]

3.2. The choice task

After agreeing to the bonus payments, and familiarizing themselves with the task, we instructed participants that we would ask them to make 20 comparisons between two plans, using dashboards that mimic 20 consecutive years of plan performance: "On each trial, the annual information on each fund is updated. You should use this information to make a decision about which of the two MySuper funds you prefer. There will be 20 trials showing 20 yearly updates, so you need to make 20 decisions in total." We then moved participants to the first of 20 choices between plans (MySuper funds) labelled XYZ and ABC (or HIJ, depending on their condition).

The experiment presented the standard dashboard information about each plan side-by-side on the screen (see Fig. 1), thereby simulating (but somewhat simplifying) actual participant comparison of competing plans. Each choice sets asked "Which of the two MySuper funds do you prefer?". Participants chose by clicking the radio button under their preferred plan. As they moved to the next set, the dashboard information updated to include the next financial year's performance for each plan, and participants chose again, thus completing 20 "years" of comparisons in total. Table 1 shows the timeline, content and number of participants of each treatment.

While we asked participants to make a new choice every period, they did not effectively face a new choice situation every period, because we progressively changed only one attribute (i.e., fees or gross returns) per treatment. We isolate the effects of either fee and gross returns information without changes to risk, and asked participants to compare two plans with the same target net returns and risk.[7] Specifically, in (fee) treatments T1 and T5, plans have the same investment strategies (and get the same gross returns), with performance differences arising solely due to fees. In (returns) treatments T2-T4; and T6-T7, similarly sized net-return performance differences arose because of improving or worsening gross returns between plans while fees stayed steady and almost equal. (Appendix D describes the net return calculations and calibrations.)

We implemented this arrangement by setting the base fees for the benchmark plan (XYZ) at the average MySuper fee for a A$50 K account balance of 1.06% (A$530 p.a.), but we varied the alternative plan fee from either a high (A$800 p.a.) or low (A$270 p.a.) starting point, consistent with the actual range of fees for these plans (see Chant et al., 2014, Table 5).[8] At each set, the alternative plan fee increased from the low starting point (or decreased

[3] https://www.nab.com.au/personal/credit-cards/calculators-and-tools/compare.

[4] We made the hold-out sample to test the effect of the performance bonus. Estimation results in section 4 show that the lack of a bonus incentive did not significantly change the plan choices of participants.

[5] All the returns presented in the dashboards are net of fees and charges (see Section 2). Since incentives were paid based on net returns, this implies that both fees and underlying gross returns were accounted for in every instance throughout the task. Furthermore, since volatility is held constant for each plan across all choices in each condition, risk is not relevant to choices or incentives.

[6] This survey length is reasonable since we test isolated changes in information, and not the full information set. In particular, the plans to choose from are essentially the same throughout each task, with one detail (fee or return) changing. The small behavioral changes between choices can thus generate the relatively short task completion time.

[7] Almost all MySuper retirement savings vehicles that operated a fixed strategic asset allocation at the time of our study held a 70:30 mix of growth:defensive assets. In 2015, 80% of fixed strategy MySuper products reported a target return above CPI of 3-4% p.a. and a "high" or "medium-high" level of investment risk (APRA 2017). A minority of MySuper funds use target date or lifecycle strategies and operate under different disclosure settings.

[8] Fees are shown for a $50 K account balance because MySuper fees consist of a fixed dollar weekly fee and a percentage fee. Investment management fees (charged

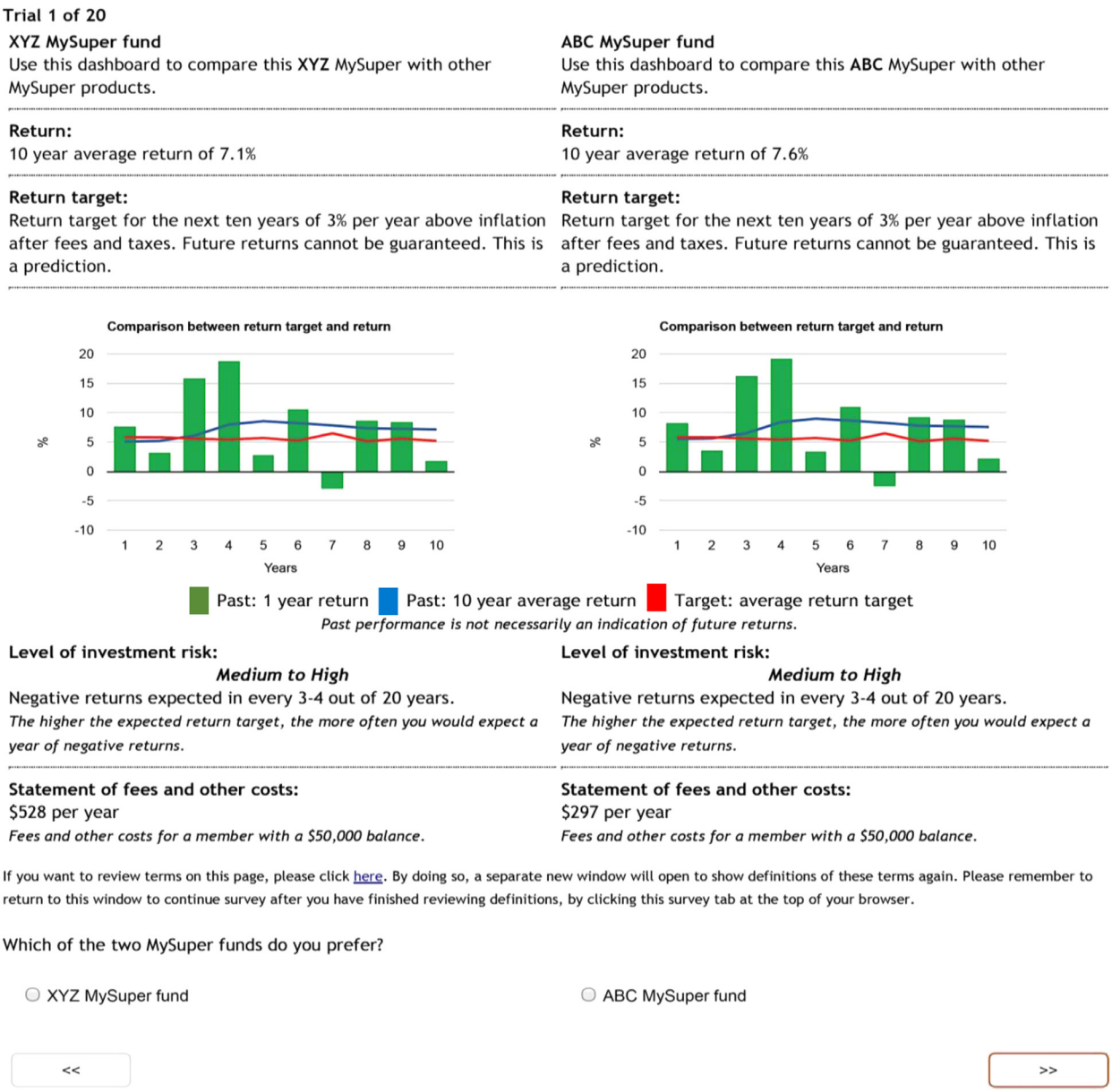

Fig. 1. Screenshot from treatment T2: Standard dashboard.

Table 1
Description of each treatment.

Treatment Number (n)	Date	Dashboard Type	Changing Information	Gross Return Volatility	Net Returns Display Format
1 (286*)	Jul 2014	Standard (Prescribed)	Fees	High	Graph
2 (274)	Sep 2014	Standard (Prescribed)	Gross returns	High	Graph
3 (252)	Feb 2015	Standard (Modified)	Gross returns	High	Table
4 (247)	Jun 2015	Standard (Modified)	Gross returns	Low	Graph/Table
5 (251)	Aug 2015	Simplified	Fees	High	N/A
6 (250)	Oct 2015	Simplified	Gross returns	High	N/A
7 (258)	Oct 2015	Simplified	Gross returns	Low	N/A

Notes: Standard (Prescribed) identifies treatments that use the MySuper dashboard format described in Commonwealth of Australia (2013) – net return target, net returns, a comparison between the net return target and the realized net returns, the projected level of investment risk, and a statement of fees and other costs for a $50 K account balance - as explained in the text. *Simplified* identifies the use of simplified dashboard format (see text for details). Variation in gross return volatility is engineered by changing the relative allocation to growth and defensive assets. In treatments T1-T3 and T5-T6 we mimic the allocation of a typical strategic asset allocation plan by including a weighted mix of growth and defensive assets, for the high volatility treatments. In treatments T4 and T7 we included only defensive assets, thus yielding a lower target return and a lower volatility of realized gross and net returns. In each treatment an approximately equal number of participants are allocated to an "increasing" and "decreasing" condition.

* 138 Incentivized and 148 Non-incentivized – all participants in remaining treatments are incentivized.

from the high starting point) by a randomly selected amount between A$20 and A$30 p.a. each year. Around the mid-point of the 20 choices sequence, the alternative plan fee equalled that on the benchmark fund and then continued to decrease (or increase) until final choice.[9] Participants who treat fees as (close to) deterministic, and who recognize that both funds have the same gross returns and are otherwise identical, should choose the low-fee plan, as distinct from some real-world settings where choosing a low fee plan might deliver less service or less skill.

In T2-T4, T6, and T7, differences in gross returns caused the differences in performance between the benchmark and alternative plans. In these treatments, fees for both plans stayed at 1.06% of a A$50K balance plus a small random adjustment at each choice set. We penalized or boosted gross returns for the alternative plan by an amount equal in dollars to the penalty (bonus) applied to plan fees in T1 and T5. The dollar value of the differences between the benchmark and alternative plans are thus the same in all treatments, but they show up either in fees (and therefore also net returns) (T1 and T5) or only in net returns (arising from gross returns, not fees) (T2-T4, T6, and T7). The regulations say that all returns in the dashboard must be net-of-fees and charges. The dashboards also always show the fees on a $50K account balance so experiment participants can infer gross returns to each plan by summing fees (translated to a percentage in the case of the standard dashboard) and net returns (from the graph or table in the standard dashboard, or from the table in the simplified dashboard).

Treatments T1 and T2 replicate the appearance of the "standard" dashboard prescribed by Australia's corporate regulator as closely as possible (see Fig. 1). Treatments T3 and T4 differ slightly from the standard dashboard by displaying net returns in a table rather than a graph (see Fig. 2). (In T4, we assigned half of participants to the table condition and half to the graph condition. In T3, we assigned all participants to the table condition.) We made this change to test the regulator's finding that participants are confused by the overlaid lines on the graph (ASIC, 2013) as opposed to findings that graphs improve comprehension and lead to better investment choices (Jarvenpaa, 1990; de Goeij et al., 2014; Kaufmann et al., 2013). Whether tables or graphs are better for showing relative returns is debateable (cf. Lohse, 1997; Vessey, 1991). Furthermore, in T4 we introduce low volatility returns by computing the gross returns of both plans from a portfolio of defensive assets. This asset allocation yields a lower target net return, lower standard risk measure, and low volatility realized net returns. The objective of this variation is to test whether our participants make different choices when relative performance signals are less volatile.

In T5-T7 we introduce a "simplified" dashboard that departs from the stipulated format but includes the same items (see Fig. 3). Specifically, we replace the graph/table of net returns history by a common percentage scale to communicate the fee, as well as the 1-year and 10-year average net returns. So rather than providing fees in dollar amounts, they are described as a percentage of a A$50K balance. We base this simplification on evidence that information presented on a common scale is more readily integrated and understood than when different metrics are used (Harries and Harvey, 2000). T5 is analogous to T1 (we vary fees), T6 to T2 (we vary gross returns) and T7 to T4 (we introduce low volatility gross returns).

By simplifying the dashboard, we can see how participants handle gross returns and fee information in different frames. We can see if measured noise declines and how the pattern of switches changes. We used the same gross returns and fees realizations (and therefore net returns) in parallel treatments in the standard and simplified dashboard treatments to allow us to measure the effect of simplification. Specifically, the participants in the fee treatments (T1 and T5) saw the same realizations of the underlying gross returns and the same fees. The subjects in the high-volatility gross-returns Treatments (T2, T3, T6) saw the same realizations of the underlying gross returns and the same fees. Subjects in the low-volatility treatments (T4 and T7) saw the same realizations of underlying gross returns and fees but from a different portfolio allocation than used to generate returns for the other treatments.

Finally, we manipulated the pattern of changes in fees and gross returns of one plan relative to the other. Aggregated studies (e.g., Sirri and Tufano, 1998) have found asymmetric responses of investors to mutual fund performance that manifests as a higher and more rapid flow of funds to outperforming managers compared with a slower movement of funds away from poor performers. Here, in each of the seven treatments there are approximately equal numbers of participants allocated to an "increasing" and a "decreasing" condition. We define the "increasing" ("decreasing") condition as the case where the net returns to the alternative plan HIJ (ABC) increase (decrease) relative to the net returns to the constant plan XYZ over the 20 choice sets. We examine these patterns to test whether people react differently to changes in relative performance due to rising versus falling net returns.

These manipulations allow us to ask how a plan participant who observes a continuous decline (improvement) in the relative net return of one plan over another should choose. When outperformance is caused by low fees, participants maximize net returns if they always choose the plan with the lowest fees. When outperformance is related to gross returns, participants may be unsure about how much is due to luck and how much is due to skill of the hypothetical investment manager. For each choice participants should still prefer the plan they judge to have the highest expected excess returns, since fees are equal and expected risk is not changing. If participants treat gross returns as a noisy signal of skill, they will gradually update their returns expectation as they learn about performance, instead of reacting instantly to news (Berk and Green, 2004). A participant who treats the gross returns signal as a noisy skill signal and observes one plan outperforming the other by a decreasing margin for several years, eventually becoming the under-performing plan, will update expectations more slowly than a participant who treats the gross returns signal as clear, and will delay longer before switching plans. So instead of assuming that participants should switch at fixed time, we *compare the average timing* of switches under the standard and simplified dashboard treatments to measure the effects of dashboard type on switching behavior.

3.3. Financial literacy and demographics

After completing the task, respondents answered questions on (i) their comprehension of the dashboard information, as well as (ii) standard inventories testing financial literacy and numeracy; (iii) pension system knowledge, and (iv) demographics, to allow us to compare our sample with the general population.[10] It is intu-

as percentages of assets) ranged from 0.32% p.a. at the 10th percentile to 0.96% p.a. at the 90th percentile and administrative fees (usually charged as a weekly nominal amount) ranged from 0.16% p.a. at the 10th percentile to 0.84% p.a. at the 90th percentile.

[9] The dashboard reports 10 years of plan performance history, via the graph or table in the standard dashboard, and via the 10- year average return, in both the standard and the simplified dashboard. To construct this 10-year history for the first choice set, we assumed that the net return difference in year 1 held for the previous 10 years. See Supplementary materials C for details.

[10] Supplementary materials D reports these additional questions via screenshots and also includes live links to all seven treatments (T1-T7), screenshots of the non-incentivized version of treatment T1, screenshots of the variations in the dashboard

Trial 1 of 20

XYZ MySuper fund
Use this dashboard to compare this **XYZ** MySuper with other MySuper products.

Return:
10 year average return of 4.1%

Return target:
Return target for the next ten years of 1% per year above inflation after fees and taxes. Future returns cannot be guaranteed. This is a prediction.

Comparison between return and target return

	Past 1 year return	Past 10 year average return	Target average return
Year 1	5.06%	4.30%	3.54%
Year 2	5.14%	4.53%	4.01%
Year 3	3.23%	4.26%	3.62%
Year 4	3.54%	4.24%	4.16%
Year 5	5.25%	4.33%	3.41%
Year 6	2.62%	4.10%	3.16%
Year 7	2.86%	3.99%	3.25%
Year 8	5.15%	4.16%	4.05%
Year 9	4.22%	4.16%	3.91%
Year 10	4.15%	4.12%	4.07%

Past performance is is not necessarily an indication of future returns.

Level of investment risk:
Very Low
Negative returns expected less than 0.5 out of 20 years.
The higher the expected return target, the more often you would expect a year of negative returns.

Statement of fees and other costs:
$523 per year
Fees and other costs for a member with a $50,000 balance.

HIJ MySuper fund
Use this dashboard to compare this **HIJ** MySuper with other MySuper products.

Return:
10 year average return of 3.6%

Return target:
Return target for the next ten years of 1% per year above inflation after fees and taxes. Future returns cannot be guaranteed. This is a prediction.

Comparison between return and target return

	Past 1 year return	Past 10 year average return	Target average return
Year 1	4.57%	3.80%	3.54%
Year 2	4.65%	4.03%	4.01%
Year 3	2.72%	3.76%	3.62%
Year 4	3.04%	3.74%	4.16%
Year 5	4.74%	3.83%	3.41%
Year 6	2.11%	3.59%	3.16%
Year 7	2.36%	3.49%	3.25%
Year 8	4.64%	3.66%	4.05%
Year 9	3.72%	3.66%	3.91%
Year 10	3.68%	3.62%	4.07%

Past performance is is not necessarily an indication of future returns.

Level of investment risk:
Very Low
Negative returns expected less than 0.5 out of 20 years.
The higher the expected return target, the more often you would expect a year of negative returns.

Statement of fees and other costs:
$536 per year
Fees and other costs for a member with a $50,000 balance.

If you want to review terms on this page, please click here. By doing so, a separate new window will open to show definitions of these terms again. Please remember to return to this window to continue survey after you have finished reviewing definitions, by clicking this survey tab at the top of your browser.

Which of the two MySuper funds do you prefer?

○ XYZ MySuper fund ○ HIJ MySuper fund

<< >>

Fig. 2. Screenshot from Treatment T4: Standard (Modified) dashboard, Table condition.

itive that, when confronted with complex information, some people are more able to handle it than others because of higher cognitive ability, more patience, better decision-making skills or previous experience. For example, earlier studies show that financially literate investors avoid higher fees (Grinblatt et al., 2015; Choi et al., 2010), but while the financially savvy tend to minimize up-front fees they do not minimize more obscure, costly expense ratios (Wilcox, 2003; Müller and Weber, 2010). Thus, we also test whether those with better comprehension of the dashboard content, more numeracy, more financial literacy and a better understanding of the retirement savings system use the dashboard differently. We incorporate general demographics including age, gender, education and income as control variables in estimation.

To sum up, the goal of our experiments is to understand responses to new information (i.e., whether the disclosed information helps people understand the difference between two plans) as the information evolves, and if this can cause them to change plans, avoiding the losses of sticking to a persistently underperforming plan. Specifically, we make systematic comparisons of different formats (graphical, tabular, simplified), dynamics (increasing, decreasing), information (fees, gross returns), and volatility (low, high), which provide comprehensive insights into the effects on participant choices of each of these features.

The design supports a range of tests of the effectiveness of the dashboard at inducing change in response to under-performance. First, we observe whether plan participants can discern and choose the dominant plan in the first "year". Second, we test how well participants understand dashboard information by whether they switched once over the 20 rounds. The monotonic path of relative performance difference over the 20 choice rounds for both fee and gross return treatments means that participants who aim for the highest expected net return on their retirement savings will choose

tested in later treatments, and a screenshot of the incentive information page. Supplementary materials E lists comprehension, financial literacy, numeracy and superannuation knowledge questions and reports summary statistics relating to participants' answers.

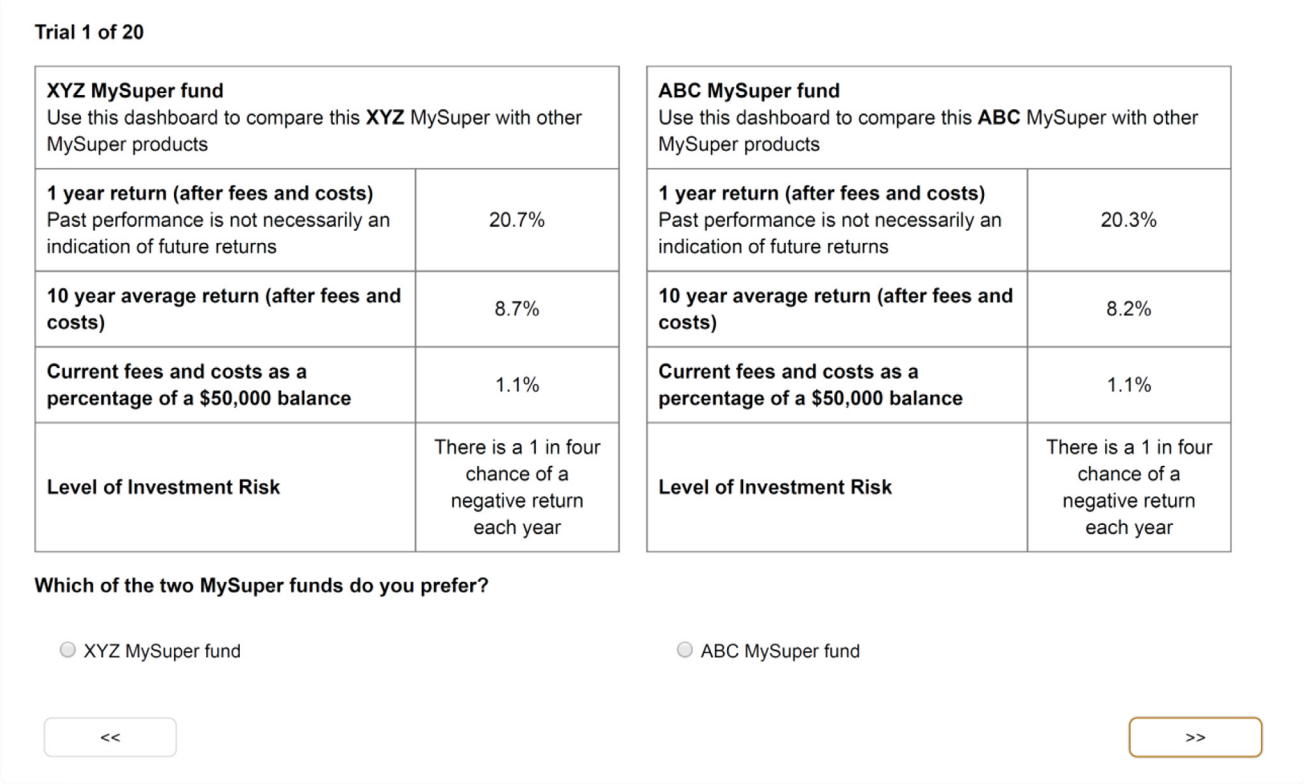

Fig. 3. Screenshot from treatment T6: Simplified dashboard.

the out-performing fund at the first set, switch once, then stay with the plan they switched to for the remainder of the 20 years of rounds. We also report differences in the speed with which participants responded to performance in different treatments and conditions. Third, we assess the relative importance participants placed on dashboard items such as net returns or fees by regressing the timing of switches on information items. And fourth, we use the timing of switches with a Bayesian model to estimate the noise that participants assign to information signals and compare these across formats, dynamics, information and volatility. We are thus able to quantify the gains associated with dashboard simplification – both in terms of improved choices and higher retirement savings. Our approach provides a richer measure of the effect of information complexity and advances earlier work on the use of heuristics (Hedesström et al., 2007; Walther, 2015), that is unincentivized (Walther, 2015) or that simply reviews participants' comments (CFPB, 2012).

4. Results

4.1. Patterns of plan choices

At the first choice-set in all treatments, one plan had a 40 basis-point higher 1-year net return and 10-year average net return than the alternative plan. Since both plan dashboards report the same expected risk and return information, and in other respects the two plans conform to the MySuper regulations, we expect participants to choose the out-performing plan at the first choice. As we expect, a large majority of experiment participants (88%) chose the plan with the highest net return at the first set. However, some of the differences between the information formats begin to emerge even at this point: around 93% of participants chose the highest net return plan in the fee treatment (T1, T5) as compared with 86% in the gross return treatments (T2-T4, T6-T7); and three percentage points more participants (90%) chose the highest net return plan in the simplified dashboard treatments than in the standard dashboard treatments (87%).

4.1.1. Single switches

We found variations across treatments and conditions in the proportions of participants who switched only once (see Table 2 Panel A). In the standard dashboard treatments (T1-T4), 75% of participants in the fee treatment (T1) switched once, but the proportion fell to below 36% when gross returns was the source of differences between plans (T2 and T3). The proportion of single switchers was higher in the low volatility returns treatment (T4) but was still less than 40%. In contrast, participants were more decisive in the simplified dashboard treatments (T5-T7), where around two thirds of the respondents in each treatment made one switch.

Table 2 Panel B reports marginal effects from logit estimations of the probability of making a single switch by treatment and in aggregate. The explanatory variables in separate treatment estimations (columns 1–7) include participants' scores for comprehension, numeracy, superannuation and financial literacy, and an indicator for passing the attention check, as well as an indicator for the "increasing" conditions. For T1, we add an indicator for participants who were not offered a bonus incentive, and for T4 we indicate the group of participants who saw the table version of the dashboard instead of the graph. Column 8 reports estimates for all treatments combined, where we include separate indicators for the simple (versus the standard) dashboard, for the treatments where performance varied because of fees (versus gross returns), for treatments where gross return volatility was set low, as well as interactions between these treatment indicators. All estimations include demographic variables as controls.[11]

[11] Demographics not separately reported in Table 2 include the following variables: *Female* equals one for female participants and zero for males; Age is a poly-

Table 2
Proportion of participants making a single switch.

Panel A: Single switching	T1 Fee	T2 Return (Graph)	T3 Return (Table)	T4 Return (Low Vol)	T5 Fee	T6 Return	T7 Return (Low Vol)	All
Single switchers (%)	76.2	21.2	35.2	39.5	70.9	70.8	64.0	54.0
Panel B: Marginal effects from logit estimations (Dependent variable: single switch =1, and 0 otherwise)								
Simple dashboard (T5–7)	–	–	–	–	–	–	–	0.244*
Fee treatments (T1, T5)	–	–	–	–	–	–	–	0.278*
Low volatility (T4, T7)	–	–	–	–	–	–	–	0.027
Increasing condition	0.019	0.275*	0.239*	0.336*	−0.020	0.035	0.096	0.120*
Simple dashboard | Fee (T1, T5)								−0.101*
Simple dashboard | Gross returns (T2–4, T6–7)								0.396*
Simple dashboard | Low vol. (T4, T7)								0.110*
Simple dashboard | High vol. (T1–3, T5–6)								0.300*
Simple dashboard | Increasing								0.146*
Simple dashboard | Decreasing								0.339*
Fee (T1, T5) | Increasing								0.195*
Fee (T1, T5) | Decreasing								0.358*
Low vol. (T4, T7) | Increasing								0.062
Low vol. (T4, T7) | Decreasing								−0.006
Comprehension	0.015	0.007	0.035	0.038	0.025	0.009	0.021	0.022
Financial literacy	0.102*	0.062	0.015	0.018	0.022	0.073	0.065	0.066*
Numeracy	0.088	−0.003	0.059	0.124	0.095	0.126*	0.160*	0.093*
Superannuation literacy	0.001	0.0004	−0.038	0.042	−0.005	0.001	0.032	−0.003
Passed attention check	0.015	0.061	0.126	0.255	0.364	0.135	−0.106	0.131*
Incentive (T1)	0.004	–	–	–	–	–	–	–
Table (T4)	–	–	–	0.053	–	–	–	–
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	286	274	253	248	251	250	258	1820
Pseudo R2	0.24	0.20	0.15	0.23	0.17	0.24	0.20	0.24

Panel A shows the proportion of participants who switch plans only once during the task. Panel B shows the marginal effects from logit estimations of the probability that a participant made one (interior) switch. Standard errors (not reported here) are clustered by participant. Effects marked * are significant at the 5% level Bonferroni-adjusted. Explanatory variables: *Simple dashboard* is a binary variable equal to one for T5-T7 and zero for T1-T4; *Fee treatment* is a binary variable equal to one for T1 and T5 and zero otherwise; *Low volatility* is a binary variable equal to one for T4 and T7 and zero otherwise; *Increasing* is a binary variable equal to 1 if the participant responded to increasing condition and zero for decreasing; *Comprehension* is the number of correctly answered comprehension questions on the dashboard; *Financial literacy* is the number of correctly answered financial literacy questions from three questions (Lusardi and Mitchell 2008); *Numeracy* is the number of correctly answered numeracy questions from three questions (Lipkus et al., 2001); *Superannuation literacy* is the number of correctly answered of 12 superannuation literacy questions (Agnew et al., 2013); *Passed attention check* is a binary variable equal to one if participants passed the attention check question that repeated an earlier question and tested whether participants recognised the repeat, and zero otherwise; *Incentive* is an indicator variable for T1 equal to one if participants were offered an incentive and zero otherwise. *Table* is an indicator variable for T4 equal to one if participants saw historical net returns information in a table and zero for a graph. In the all-treatment estimation, we also include interactions between the indicator *dashboard* and indicators for *Fee treatment, Low volatility* and *Increasing*, plus interactions between the indicator for *Increasing* and indicators for *Fee treatment* and *Low volatility.* Demographics defined in footnote 11 are included as controls in all estimations.

Single switches were 24 percentage points (ppts henceforth) more probable when participants saw the simplified instead of the standard dashboard, and 27 ppts more probable in the treatments where plan fees, instead of gross returns, varied. Moreover, lowering the volatility of gross returns did not make a significant difference. For the gross-returns treatments, viewing the simple instead of the standard dashboard raised the probability of a single switch by nearly 40 ppts, while the effect of simplification was much less (11 ppts) for the low-volatility treatments (column 8, rows 6 and 7). By contrast, for the fee treatments, switching from the standard to the simplified dashboard, significantly *reduced* the probability of single switches by 10 ppts, the effect of "simplification" working to obscure, rather than clarify fees (column 8, row 5).

Another framing effect is worth noting: a single switch was 12 percentage points (ppts) *more* likely for "increasing" conditions. Indecision by participants in decreasing conditions resembles the mutual fund investors' tendency to withdraw from poorly performing funds less readily than they get into highly performing ones (Sirri and Tufano, 1998). This tendency, however, is also related to the information format: participants who were allocated to the increasing conditions in standard dashboard return treatments T2-T4 were significantly more likely to make only one switch than those in the decreasing conditions. In other words, participants had more difficulty evaluating performance variations due to decreasing rather than increasing gross returns when using the standard dashboard. Specifically, participants in the increasing condition were 15 ppts more likely to switch once in the simplified, than in the standard, dashboard setting, but participants in the decreasing conditions were 34 ppts more likely to switch once when viewing the simple rather than the standard dashboard. We conclude that the simplified dashboard helped participants to notice changes in the decreasing gross-returns conditions.

Participants who scored high on tests of comprehension, financial literacy and numeracy, and participants who passed the attention check, were significantly more likely to switch only once. We do not detect a relevant difference in the proportion of single-switchers in the non-incentivized group in T1, or in the group who

chotomous variable equals 0 if participants are under 39 years old, 1 if between 40 – 59 years old, and 2 if over 60 years; *Married/de facto* equals one if married or living in de facto relationship and zero otherwise; *Financial decision* equals one if the participant himself/herself is most responsible for the major financial decisions and zero otherwise; *No dependents* equals one if the participant only supports himself/herself financially and zero if more than one person; *High school graduate* equals one if the participant graduated from high school and zero otherwise; *College diploma/degree* equals one if the highest school qualification is Bachelor Degree/Graduate Diploma/Master Degree/PhD and zero otherwise; *Employed* is a polychotomous variable taking the value zero if unemployed/not in the labour force (inc. stay-at-home parents, full-time students, or others), one if employed part-time or full-time, and two if retired; *Weekly income* is a polychotomous variable taking the value of zero if negative or nil weekly (annual) gross personal income (before tax), 1 if A$1-A$399 (A$1-A$20,799), 2 if A$400-A$999(A$20,800-A$51,999), 3 if A$1,000 or more (A$52,000 or more); and *Retirement balance* is the log of participants' reported retirement account balance or zero for missing balance.

viewed the table instead of the graph of historical net returns in T4.

4.1.2. Timing of switches

Most fee treatment (T1, T5) participants switched at or very near the choice set where the outperforming plan actually changed (i.e., around choice set 11 or 12). In gross returns treatments (T2-T4, T6-T7), however, participants delayed switching until well after the change in the outperforming plan. Table 3 records the pattern of switches by choice set, treatment and condition for the standard (Panel A) and simplified (Panel B) dashboard treatments. The rows correspond to the choice sets numbered from 2 to 20, while the columns correspond to treatments, divided into separate counts of (i) single switches ("single"), (ii) first switches ("first"), and (iii) final switches ("final"). Each of these columns is further split to reflect the associated increasing ("I") or decreasing ("D") conditions. Hence, every cell in Table 3 shows the number of participants in that condition who made their only, first or final switch at that choice set. The dark gray cells show the choice set where the outperforming plan changed in each treatment; the pale gray cells show the choice sets where the *10-year average return* information is either equal between plans or clearly higher (lower) for the alternative plan. So, if participants choose only on the basis of higher *10-year average net returns* they will switch at the first pale gray shaded cell. For the standard dashboard we find most participants in the *fee* treatment (T1) choosing to switch plans at, or immediately after, the point at which the outperforming (lowest fee) plan changed. In other words, most participants chose the plan with the smaller nominal dollar fee. Many people in gross return treatments (T2-T4), however, delayed switching, most switching back and forward between plans several times. In high volatility gross returns treatments (T2-T3), for instance, the majority of participants waited at least 3–6 sets after the outperforming plan switched, and many continued to switch between the alternatives until the end of the 20 rounds. Only two respondents out of 499 in T2 and T3 'chased' the highest net return at each choice set.

Turning to the simplified dashboard (Table 3 Panel B) we find: first, "simplifying" the fee information - by expressing it as a percentage rather than in absolute dollars - presumably makes the fees less salient and harder to evaluate.[12] In the simplified fee treatment (T5) participants wait longer than for the standard dashboard before leaving the higher fee plan. Also, the simplified net returns information competes more strongly for people's attention in the simplified than in the standard dashboard, although performance differences are really caused by fees. There are groups in T5 who wait until around the 15th choice round to switch, suggesting they respond to differences in 10-year average net returns rather than fees. Overall, those who see the simplified fee dashboard delayed switching to the lowest fee fund more often. We thus conclude that the "fee simplifications" we introduce do not help participants make better choices.

In contrast, when we consider gross returns (T6-T7), we see clear signs that the simplified format changes choices. Many more people switch only once between plans, and although they still wait a few rounds before they commit to the higher-gross-returns plan, the delays are typically shorter than for the equivalent standard dashboard treatments. Such patterns of more confident switches hold for both high and low gross return volatility treatments. This implies that the way the standard dashboard frames returns is considerably more confusing to participants than in the simplified format.

[12] Our result is largely consistent with studies of mutual fund investors using market and experimental data that find that investors are less sensitive to percentage fees (operating expense fees) than to front end loads (e.g., Barber at al. 2005; Anufriev, Bao et al. 2019).

4.1.3. Effects of dashboard information items on plan switches

We now report the effects of specific dashboard information items on plan choices. Table 4 reports the average marginal effects from panel logit models of first switches for each of the treatments (T1-T7) by increasing/decreasing conditions. (Similar models of final switches show consistent results.) We construct the dependent variable to indicate the first point in the sequence of 20 choices where a participant responds to the gradually increasing differences in plan performance. For increasing conditions, the dependent variable is an indicator variable that equals one while the participant chooses XYZ (the left-hand side plan) and changes to zero when the participant first chooses the alternative plan. For decreasing conditions, the dependent variable equals zero until the participants first chooses XYZ (the left-hand side plan) and takes the value of one from then on. We define three information variables: Δ *1 yr ret* is the difference between the 1-year net return to plan XYZ and the 1-year net return to plan ABC or HIJ (on the right-hand side). Similar definitions apply to the differences in the 10-year average net return (Δ *10 yr ret*) and the difference in fees (Δ *Fee*). We expect positive differences in returns (higher values of Δ *1 yr ret* and Δ *10 yr ret*) to increase the probability of choosing XYZ and the reverse response for fees. We also include an indicator variable that takes the value of one if the participant makes only one switch in the sequence and zero otherwise. For the standard dashboard models (T1-T4), we find the marginal effects of Δ *Fee* from T1 models to have the expected negative sign, and indicate a 20% lower chance of participants choosing the XYZ plan for fees A$100 p.a. higher than the alternative. However, results from T2-T4 show that when similar differences in performance show up in gross returns (rather than fees), participants do not react (i.e., all the marginal effects of Δ *1 yr ret* are not significant). And reducing the volatility of gross returns in T4 does not change this outcome. In contrast, a higher Δ *10 yr ret* makes first switches to XYZ more likely for the decreasing condition of T2. In this case, a 0.5% p.a. higher 10-year average net return, for example, makes first switches to XYZ 35% more likely. This result suggests that 10-year average returns are easier to see and possibly judged as more reliable signals of performance. Participants also seem to note the fee differences in T2-T4, even though the variations are small and randomized. In T4, the graph is also related to slower switches than the table. Finally, in decreasing conditions, participants who switched once, waited longer to do so.

For the simplified dashboard, the Δ *10 yr ret* is a significant and positive predictor of first switches for all but two conditions. The Δ *1 yr ret* also influences choices in two more cases than in the standard dashboard, with the expected positive sign. Higher fees deter participants viewing the simplified dashboard from switching to XYZ, as expected. (Differences in marginal effect sizes of Δ *Fee* between the upper and lower panel of Table 4 are due to the switch to percentage fees in the simplified dashboard.) To sum up, Tables 3 and 4 show that participants interpret fee and gross return information in a way we might expect: they prefer low fee, high net returns plans. However, while large changes in fees prompt people to switch almost immediately, it takes more and larger gross returns signals to prompt change. Participants take notice of both short-term and long-term net returns, but delay switching to the better performing plan until after they have seen several years of short-term outperformance. Even when volatility is low, participants still wait to switch. Comparing the standard and simplified dashboards shows that participants' delay to switch in response to relative gross returns is not only due to a cautious appraisal of noisy returns signals (a view probably reinforced by the warnings placed on both formats), but also due to the additional, and presumably confusing, information in the standard dashboard.

These results give a nuanced interpretation of how people might use past performance information. While participants com-

Table 3
Numbers of participants switching at each choice set.

Panel A: Standard dashboard

	Fee (T1)						Gross Returns/Graph (T2)						Gross Returns/Table (T3)						Gross Returns/Low vol (T4)					
Switch	**Single**		**First**		**Final**		**Single**		**First**		**Final**		**Single**		**First**		**Final**		**Single**		**First**		**Final**	
	I	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**
Set																								
2			5	6	1				8	12		2	1	1	5	7	1	1		1	5	7		1
3			3	2					3	3					7	6	1				5	4		1
4			1	2					2	2						2					11	2		
5			2	2					1	3					1	2	1				2	4		
6				2					3	2					1	1					2	4		
7			1	2					2	5	1				1	4								
8	2		2	1	2					1						1					3	2		
9				2						3						3						3		
10	1	7	2	7	1	10											1					4	1	
11	2	52	3	56	4	56			3			1										7		
12	29	17	32	21	30	18				1			2		2		2	1			2			
13	34	21	34	21	37	25			5	6		2			5	1	2			1	1	3	3	3
14	19	6	19	6	23	14				56			1	2	2	45	1	7	11	5	15	40	16	6
15	16	10	16	12	19	10	4		12		9		29	4	46	5	38	5	8	2	8	3	12	6
16		1		1	3	3	37		51		51	1	11	1	14	1	12	5	40	6	45	9	44	40
17	1		1		3	3	3	8	4	12	6	81	4	14	4	15	6	47	5	5	5	5	8	32
18					2	3	2		2	1	9	6	7		7		12	4	3	2	5	2	8	7
19					5	4	1	2	1	4	36	34	3	4	3	5	28	34	2	4	2	4	9	14
20					4						19	4	3	2	3	2	15	10		1		1	14	5
% Loss	I: -0.26*			D: -0.29*			I: -0.98			D: -1.20			I: -0.82			D: -1.30			I: -0.78			D: -0.95	

Panel B: Simplified dashboard

	Fee (T5)						Gross Returns; (T6)						Gross Returns; Low vol (T7)					
Switch	**Single**		**First**		**Final**		**Single**		**First**		**Final**		**Single**		**First**		**Final**	
	I	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**	**I**	**D**
Set no.																		
2			7	3					4	3					9	4		
3			1	1					2	6					3	3		
4			2	4						4					7	1		
5			1	1						3							1	
6				2	3				4	1					2			
7			1		1				2	1						1		
8										1	1							
9									1	1							1	1
10		1	4	3		1				1	1	1		1		11	1	1
11	1		2	2	1	1		2		6		5				9	1	
12	29	13	36	15	35	13						1	9		12		12	
13	16	44	17	52	18	47	6	6	9	6	9	7	1	14	5	14	2	26
14	13	9	13	11	17	9	10	70	10	71	11	76		56		59	2	72
15	18	24	18	24	23	38	60	2	62	3	68	8	10	5	11	5	15	9
16	1	1	1	1	3	5	6		6		8	4	49	2	49	2	62	7
17	1	1	1	1	4	2	1	8	1	8	5	12	12	2	12	2	15	3
18	2	2	2	2	6	5	3	1	3	1	7	4		1		1		7
19					2	5					4	2					3	4
20					3	1		1		1	8	5					2	
% Loss	I: -0.40				D: -0.45		I: -0.59*				D: -0.75*		I: -0.81				D: -0.49*	

Notes: This table shows the number of participants that switch plans at each choice set. The "Single" column shows participants who made one switch in 20 choices; the "First" column shows the first switching point for all participants who made one or more switches; the "Final" column shows the last switching point for all participants who made one or more switches. The "I" indicates conditions where the net returns to alternative plan (HIJ) are increasing relative to the constant plan (XYZ); "D" indicates conditions where the alternative plan net returns (ABC) are decreasing relative to the constant plan (XYZ). The dark gray cells show cross-over points where the out-performance changes from one plan to the other. The pale gray cells show sets where the 10-year average net returns to HIJ (ABC) are equal to or higher (lower) than (XYZ). The last row shows the average percent of balance lost to mis-timed switching and the result of a *t*-test for equality of losses between standard and simplified dashboard treatments. $^{*}p<0.05$.

Table 4
Marginal effects of information variables on plan switches.

	First switch						
	Δ1 yr ret	Δ10 yr ret	ΔFee	Single	Graph	Ps. R^2	Obs.
Standard dashboard							
T1 (FEE, GRAPH)							
Increasing			−0.002***	−0.024		0.433	2780
			0.000	*0.070*			
Decreasing			−0.002***	−0.156***		0.595	2940
			0.000	*0.052*			
T2 (GROSS RETURNS, GRAPH)							
Increasing	0.863***	−0.061	0.001	0.058		0.232	2720
	0.105	*0.114*	*0.001*	*0.037*			
Decreasing	0.269	0.727**	−0.005***	−0.228***		0.267	2760
	0.245	*0.283*	*0.001*	*0.034*			
T3 (GROSS RETURNS, TABLE)							
Increasing	0.473	0.440	−0.004*	0.055		0.287	2520
	0.278	*0.315*	*0.001*	*0.044*			
Decreasing	0.258	0.693*	−0.004***	−0.149***		0.261	2520
	0.255	*0.293*	*0.001*	*0.046*			
T4 (LOW VOLATILITY RETURNS, GRAPH or TABLE)							
Increasing	0.187	0.751	−0.005***	0.250***	0.148***	0.395	2460
	0.293	*0.331*	*0.001*	*0.044*	*0.042*		
Decreasing	0.193	0.805*	0.003***	−0.180***	−0.039	0.315	2480
	0.280	*0.318*	*0.001*	*0.047*	*0.047*		
Simplified dashboard							
T5 (FEE)							
Increasing		0.505***	−0.556***	0.073		0.423	2400
		0.081	*0.063*	*0.059*			
Decreasing		−0.313***	−1.270***	−0.101		0.518	2620
		0.079	*0.061*	*0.054*			
T6 (GROSS RETURNS)							
Increasing	0.984***	−0.020		0.032		0.414	2460
	0.041	*0.046*		*0.062*			
Decreasing	−0.019	1.189***	−1.843***	−0.215***		0.481	2540
	0.088	*0.100*	*0.220*	*0.057*			
T7 (LOW VOLATILITY RETURNS)							
Increasing	−0.041	1.030***	−1.399***	0.265***		0.413	2460
	0.068	*0.086*	*0.095*	*0.055*			
Decreasing	0.298***	0.711***	0.196	−0.015		0.419	2700
	0.043	*0.041*	*0.152*	*0.047*			

Notes: This table shows the estimated marginal effects of explanatory variables from logit models of participants' first switches in 20 plan choices. The "Δ1 yr ret" is the difference in the 1-year net returns (XYZ-ABC/HIJ); "Δ10 yr ret" is the difference in the average 10-year net returns; "ΔFee" is the difference in fees; "Single" is a binary indicator for participants who made one switch between funds; "Graph" is a binary indicator for when historical returns are presented as a graph (not a table). Variables are omitted from models of T1 and T6 (increasing) because of collinearity. Standard errors are clustered by participant. The delta-method standard errors in italics.

* $p<0.1$;.

** $p<0.05$;.

*** $p<0.01$.

pare *10-year average net returns*, far fewer used the *1-year net return* regardless of whether it was shown in a graph or table. This focus on long-term average returns is consistent with the experimental work of Wilcox (2003) and the evaluation of aggregate revealed preference data (Benartzi, 2001; Benartzi and Thaler, 1999; Choi et al., 2010), but is somewhat at odds with empirical studies on aggregated mutual fund flows that find investors choose funds with strong recent performance (Sirri and Tufano, 1998; Del Guercio and Tkac, 2002; Frazzini and Lamont, 2008; but see Anufriev et al., 2019). This behavior could also be related to the warnings' position in the standard dashboard – next to the graph but away from the *10-year average return*. In the simplified dashboard, the *1-year return* is much easier to see, but still positioned next to a warning against projecting future returns from past performance, while the *10-year average return* is not.

The relatively simple fee presentation in the standard dashboard can also explain why our participants generally treat fee information as highly useful. Moreover, the fact that they give more weight to dollar-based fees in the standard dashboard – rather than comparable percentages in the simplified dashboard – confirms previous findings (Wilcox, 2003; Barber et al., 2005; Beshears et al., 2011; Choi et al., 2010). Our results go further than those in previous studies, by showing this to be a pure framing effect.

4.2. A Bayesian estimate of the effects of information complexity

The dashboard warns people that past returns are not necessarily an indication of future performance, so participants might be sceptical of gross returns differences between plans since they do not know their source, or they could treat gross returns as a noisy signal of manager skill. We address this issue by comparing participants' delayed switches across different treatments using a Bayesian updating model with normal priors. The model estimates the relative noisiness of the signals from fees and gross returns across different treatments. The differences between these estimates for the standard and simplified dashboards measure the effect of information framing. Importantly, note that we do not propose that participants actually use this exact process when deciding on plans, even though we observe people choosing in a qualitatively similar way. We merely treat choices as if the model is true and use it to measure and compare signals.

Consider a plan participant who believes the 1-year net rate of return to plan i (XYZ; ABC, HIJ) is a normally distributed random variable $Y_i,\ i=1,\ 2$ consisting of the true signal of the plan's per-

formance x_i and a noise component ε, related to unpredictability and assumed the same for both plans:

$$Y_i = x_i + \varepsilon, \text{ where } x_i \sim N(\mu_{x_i}, \sigma^2_{x_i}), \ \varepsilon \sim N(0, \sigma^2_\varepsilon)$$
$$\text{and } \varepsilon \text{ is independent of } x_i.$$

At the first choice set , participants use the return target and risk information to choose parameters for the prior distribution of x_i. The two plans are designed to have the same strategic asset allocation and thus the same return target and risk. Thus, by assumption, a person choosing between them will hold the same prior distributions for both plans. Since both versions of the dashboard show the return target as a real rate of return over the CPI, a participant must estimate future inflation to calculate (nominal) μ_x. We use the Australian central bank's inflation target band midpoint (2.5% p.a.) as the inflation expectation. Next, for σ^2_x, the plan participant must combine her μ_x estimate with the quantile information in the dashboard risk section. In T1-T3 and T5-T6, a return target of CPI+3% $\mu_x = 0.055$ and risk information stating that the probability of a (nominal) return below zero is 3–4 years in every 20 (Pr(x<0)=0.175) means a normal standard deviation of 5.9% p.a. (σ^2_x =0.003481). In T4 and T7 (low noise treatments), these parameters are $\mu_x = 0.035$ and σ^2_x= 1.8% p.a. The participant then forms a posterior distribution for x_i by updating her beliefs about its mean at each new choice set. (We assume a constant estimate of the variance σ^2_x, as the target risk measure remains constant across rounds and between plans.) Participants update their posterior mean using the new signal about 1-year net returns for each plan, $y_{i,t}, \ t = 1, 2 \ldots 20$, weighted by their beliefs about the relative variability of the true distribution of x and ε. The posterior mean at each choice is:

$$E_t[x_i|Y_i = y_{i,t}] = \mu_{x_i} + \frac{\sigma^2_x}{\sigma^2_x + \sigma^2_\varepsilon}(y_{i,t} - \mu_{x_i}).$$

Starting with a prior expectation for x based on the return target and risk information at the first choice set, participants compute the posterior $E_1[x_i|Y_i = y_{i,1}]$ for each plan conditional on the 1-year net return signal, and use this posterior expectation as the prior value of μ_{x_i} for the second choice set, and so on. At each choice set, participants should prefer the plan with the higher (posterior) expected net return $E_t[x_i|Y = y_{i,t}]$. If $\sigma^2_\varepsilon = 0$, then $E_t[x_i|Y = y_{i,t}] = y_{i,t}$; participants will prefer the plan with the higher net return and switch when the dominant plan changes. But as σ^2_ε increases, the posterior mean changes more slowly and people will wait for more evidence of superior performance before switching.

Using this setup, we can infer a value for $\hat{\sigma}_\varepsilon$ that justifies a switch at each choice set after the set at which the outperforming plan changes. We can assign this value to participants who delay their switch to any choice set. To make the comparison easier, we report the inferred value of $\hat{\sigma}_\varepsilon$ as a scaling of σ_x so that $\hat{\sigma}_\varepsilon = \hat{s}\sigma_x$. The scaling factor indicates the proportions of signal and noise that participants extract from the net returns in this updating model (and thus can be estimated for each participant). This scaling is an evaluation instrument for the scepticism that people attribute to returns in different treatments. Since the underlying gross return and fee information is the same across equivalent treatments in the standard and simplified dashboard (e.g., T1 versus T5, T2 versus T6, T4 versus T7) - then we can attribute the differences in $\hat{S}$ estimates between these pairs to information framing, i.e., to simplification.

Table 5 (Panel A) reports the treatment-specific means for the $\hat{S}$ factor that the participants' final switches indicate. For the standard dashboard treatments, the mean scaling factor for the fee treatment (T1) is less than half the means for the gross returns treatments, including the low volatility gross return treatment (T2-T4). In other words, participants treat gross returns signals as roughly twice as noisy as fee signals when delivered through the standard dashboard.

Similarly, when we compare the mean scaling factors of the gross returns, standard dashboard treatments with the simplified dashboard treatments, we see a large reduction: the mean scaling factor for participants in T2 was 4.10, compared with 2.54 for participants in T6. It follows that a major reason for later switching in the standard dashboard gross returns treatments is due to its complex format, over and above any uncertainty about whether gross returns are themselves reliable signals of plan performance. But dashboard simplification also *raises* scaling factors for T5 relative to T1. Changing fee information – from nominal dollars to percentage – appeared to add to the confusion. Our results thus do not support our expectation that people would benefit from seeing fees and returns in a scale compatible way.

Table 5 (Panel B) reports marginal effects from OLS regressions of participants' scaling factors on indicators for design features and participants' various literacy and attention scores (as defined in the notes to Table 2) for T1-T7 in columns 1–7. Column 8 reports estimates for all treatments combined, where we include indicators for the simple (versus the standard) dashboard, for the fee (versus gross-return) treatments, for low volatility gross-returns treatments, as well as interactions between these treatment indicators. All estimations include demographic variables as controls.

Going from the standard to the simplified dashboard significantly lowers scaling factors by 0.753 overall (column 8, row 1), indicating a higher average clarity for the simplified dashboard. This average simplification effect, however, combines a significantly *higher* (0.955; column 8, row 5) scaling factor for the simplified fee treatment (T7) with significantly *lower* factors for the gross returns treatments (−1.468; column 8, row 6). Lower volatility on its own, in either simple or standard dashboards, is associated with lower scaling factors, but the estimated effect is not statistically significant after Bonferroni adjustment. By contrast, the marginal effect of fee treatment versus gross returns treatment is to lower scaling factors by 1.643.

Consistent with our analysis of single switches (Table 2), these estimates show that participants responded more slowly to performance variations due to decreasing, rather than increasing, gross returns when using the standard dashboard. The marginal effect from the increasing condition indicator in T2-T4 is to lower scaling factors by between 1.5 and 0.76. but is confined to the standard dashboard for gross returns treatments. We also find a similar-sized marginal effect for the simplified dashboard fee condition (T5).[13] Participants who scored high on numeracy also responded significantly more quickly to signals related to fees and gross returns.

To sum up, these results show that participants respond to fee signals generally more promptly than to gross returns signals. A major cause of the slower speed of response to gross returns signals is the standard dashboard format itself, but the effect of the confusing format is worse for participants assigned to the decreasing condition. At the same time, participants respond to fee differences significantly more slowly in the simplified dashboard, that shows percentage fees, than in the standard dashboard that shows

[13] The results in column 5 show that the indicator for the increasing/decreasing condition is significant and negative. This result is different from the T5 logit model of the probability of switching only once, reported in Table 2, where the "increasing" indicator was insignificant. The scaling factors used as the dependent variable in these regressions are calculated from the timing of the last switch that all participants make. It follows that more confused participants who switch more than once and at a later choice set, will influence the scaling factor regressions. The difference between these results can thus be explained by the difference between the dependent variables and by the observation that the decreasing condition is associated with more confusion, and that the percentage fees are harder for participants to understand.

Table 5
Scaling factors.

Panel A: Scaling factors	T1 Fee	T2 Return (Graph)	T3 Return (Table)	T4 Return (Low Vol)	T5 Fee	T6 Return	T7 Return (Low Vol)	All
Mean	1.54	4.10	4.20	3.38	2.27	2.54	2.32	2.89
Panel B: Marginal effects from OLS estimations (Dependent variable: participant scaling factor)								
Simple dashboard (T5–7)	–	–	–	–	–	–	–	−0.753*
Fee treatments (T1, T5)	–	–	–	–	–	–	–	−1.643*
Low volatility (T4, T7)	–	–	–	–	–	–	–	−0.507
Increasing condition	0.440	−0.763*	−1.519*	−1.421*	−1.010*	−0.132	−0.564	−0.676*
Simple dashboard | Fee (T1, T5)								0.955*
Simple dashboard | Gross returns (T2–4, T6–7)								−1.468*
Simple dashboard | Low vol. (T4, T7)								−0.395
Simple dashboard | High vol. (T1–3, T5–6)								−0.890
Simple dashboard | Increasing								−0.655
Simple dashboard | Decreasing								−0.847
Fee (T1, T5) | Increasing								−1.384
Fee (T1, T5) | Decreasing								−1.891
Low vol. (T4, T7) | Increasing								−0.641
Low vol. (T4, T7) | Decreasing								−0.379
Comprehension	0.009	0.008	−0.050	0.062	0.006	−0.038	−0.113	0.028
Financial literacy	−0.469	−0.019	−0.049	−0.439	−0.283	−0.303	0.236	−0.198
Numeracy	−0.262	−0.195	−0.404	−0.729*	−0.434	−0.110	−0.252	−0.322*
Superannuation literacy	0.068	−0.166	0.260	−0.695	−0.367	0.128	−0.546	−0.210
Passed attention check	−1.059	0.353	−0.188	0.452	−0.610	0.213	0.031	−0.210
Incentive (T1)	0.096	–	–	–	–	–	–	–
Table (T4)	–	–	–	0.572	–	–	–	–
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	286	274	253	248	251	250	258	1820
R^2	0.24	0.16	0.23	0.29	0.21	0.10	0.15	0.26

This table shows the mean (Panel A) and OLS regression estimates (Panel B) of Bayesian scaling factors for the standard deviation of noise distribution. Panel B reports the marginal effects from OLS regressions of scaling factors on indicators for experimental treatments and participant characteristics. We also include interactions between the indicator *Dashboard* and indicators for *Fee treatment, Low volatility* and *Increasing*, plus interactions between the indicator for *Increasing* and indicators for *Fee treatment* and *Low volatility.* Demographics as defined in footnote 11 are included as controls in all estimations. The explanatory variables are defined in notes to Table 2. The estimates marked with an asterisk are significant at the 5% level or less after Bonferroni adjustment. A Bayesian scaling factor of two indicates that the standard deviation of the participant's individual noise distribution is twice as large as the standard deviation of the underlying net return distribution. Scaling factors are assigned to participants in each treatment according to the choice set at which they make a final switch with larger scaling factors corresponding to later switches. We assign a scaling factor of zero to the participants who make their last switch at or before the set at which the outperforming plan changes.

fees in dollars. Again, this slower response was significantly increased for participants in the decreasing condition.

4.3. Dashboard complexity: the indicative cost

An inefficient choice of pension plan will be costly to plan participants. To estimate how costly it would be, we compute an indicative final account balance after 20 "years" for each participant. We assume that they begin with a A$50K balance and do not contribute or withdraw any savings. (In the gross returns treatments, for example, average final account balances are around A$155K.) Our account balance calculation includes deductions for fees and charges. We also compute the maximum possible final account balance and calculate the average percentage difference between participants' realized and maximum balances. The last row in each panel in Table 3 reports the average (per participant) percentage loss for each treatment and condition. We then test for the equality of the average percentage losses between the standard and simplified dashboard treatments. If differences are significant then less complexity makes participants better off, on average. Two factors significantly affect the average losses we calculate here – the source of the outperformance signal and information framing. Average losses are lowest (0.3%) in the standard dashboard fee treatments but are three or four times as large in the standard dashboard gross returns treatments (up to 1.3% of final account balance).

Hence, some basic format simplifications make a considerable difference to losses. Participants in gross returns treatments who see the simplified dashboard incur losses on average well below average losses incurred by participants who see the standard dashboard, leading to differences of up to 0.5% of their final account balance. However, reframing nominal fee information into annual percentages did not help, increasing losses by around 0.1%.

5. Conclusions

Many studies document firms' use of complex financial products and price information to preserve price dispersion, limit competition and hinder optimal choice (Carlin, 2009; Carlin and Manso, 2010; Henderson and Pearson, 2011; Célérier and Vallée, 2017). For retirement savings decisions, the issue is exacerbated by people's confusion with such choices, and the consequent minimal competitive discipline they apply to pension plan providers. While it can be argued that plan participants can simply read plan disclosures to inform themselves, even standardized summary product disclosures can have unforeseen effects on participants' decisions.

We evaluated a new summary information format, called a dashboard, intended to make it easier for retirement savers to compare pension plans. We tested this dashboard first in its standard form (as prescribed by the Australian regulator in 2014), and then with various "simplifications" in a sequence of choices that represents a plan participant's working life accumulation. The dashboard was designed to make plan comparisons easier, in the expectation that it would over-ride the tendency of plan members to stick with under-performing plans, a well-documented and costly problem in the Australian retirement savings sector.

We advance understanding beyond earlier studies of complex disclosures and financial expertise by (i) separately identifying the effects of simplified fee and gross return disclosures on plan

choices; (ii) estimating participants' responses to sequential performance information in different formats; and (iii) assigning a magnitude and a value to the reduction in information complexity associated with different information formats. In doing so, we also provide a timely contribution to the current policy debate advocating direct comparisons of common financial products.

First, and in some contrast to previous research (Wilcox, 2003; Choi et al., 2010; Beshears et al., 2011), we find that people react quickly when outperformance shows up in nominal fee differences. When fees are percentage-based, however, they switch away from the under-performing plan at a considerably slower pace. Our dashboard simplifications, which aimed to help people integrate fee and return information by presenting both as percentages, actually made fees less "obvious", thus supporting the disclosure of fees in nominal dollars.

Second, we infer that a major reason for the slow, disordered switches people make when they look at the standard dashboard is the confusion caused by its complex format (over and above any uncertainty about the past performance – manager-skill link). The simplified dashboard enables people to switch away from under-performing plans faster, conditioning on the beliefs about the signal value of gross returns. Indeed, simplifying net returns information reduces the noise participants assign to gross returns by about 40%. We test whether this delay is due to "volatility aversion", but see that even with very low return volatility, dashboard simplification affects choice patterns, and noise estimates still drop by more than 40%. Overall, information complexity, even presented in a standard dashboard format, makes performance evaluation difficult, so that better comprehension can be achieved by improving information formats, resulting in significant benefits for retirement savings.

Third, we show that identifying the right disclosure simplifications is not straightforward. Firms or regulators that want to improve information disclosures should first find out what and how information is used, before designing new formats. Techniques that might be expected to improve comprehension can be ineffective. The methods regulators used to test the dashboard we study here – focus groups and interviews testing comprehension – are standard internationally. But our results support previous views that such methods are insufficient when it comes to informing product design and policy (Gillis, 2015; Bateman et al., 2016). Our incentive-compatible experimental testing reveals the importance of going beyond such methods when assessing the comprehension, use and effectiveness of alternative disclosure formats.

Funding

This work was supported by the Australian Research Council [LP110100489], the Australian Research Council Future Fellowship [FT110100151] (Newell) and the ARC Center of Excellence in Population Ageing Research [CE110001029] (Dobrescu).

CRediT authorship contribution statement

S. Thorp: Conceptualization, Methodology, Formal analysis, Validation, Writing - original draft, Writing - review & editing, Funding acquisition. **H. Bateman:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Funding acquisition. **L.I. Dobrescu:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Funding acquisition. **B.R. Newell:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Funding acquisition. **A. Ortmann:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Funding acquisition.

Acknowledgements

We are grateful to Sean Anthonisz, Doug Foster, Rawley Heimer, staff at UniSuper, participants at the Research in Behavioural Finance Conference, VU Amsterdam 2016, the Colloquium of Superannuation Researchers UNSW Sydney 2016, the Econometric Society Australasian Meeting, UTS Sydney 2016, the Netspar International Pension Workshop, Leiden 2017 and seminar participants at the School of Economics, University of Sydney, the Department of Finance, University of Melbourne, and the Department of Applied Finance, Macquarie University, for helpful comments. We received excellent research assistance from Ang Li, Sol Chung and Jiahuan Zhang. Funding was provided by Australian Research Council LP110100489.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jbankfin.2020.105955.

References

Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S., Evanoff, D., 2014. The effectiveness of mandatory mortgage counselling: Can one dissuade borrowers from choosing risky mortgages? National Bureau of Economic Research Working Paper w19920. NBER, Cambridge.
Agnew, J.R., Bateman, H., Thorp, S., 2013. Financial Literacy and Retirement Planning in Australia. Numeracy 6 (2) Article 7.
Anufriev, M., Bao, T., Tuinstra, J., 2016. Microfoundations for switching behavior in heterogeneous agent models: an experiment. J Econ Behav Organ 129, 74–99.
Anufriev, M., Bao, T., Sutan, A., Tuinstra, J., 2019a. Fee structure and mutual fund choice: an experiment. J Econ Behav Organ 158, 449–474.
Anufriev, M., Chernulich, A., Tuinstra, J., 2018. A laboratory experiment on the heuristic switching model. J Econ Dyn Control 91, 21–42.
Anufriev, M., Hommes, C., Makarewicz, T., 2019b. Simple forecasting heuristics that make us smart: evidence from different market experiments. J Eur Econ Assoc 17 (5), 1538–1584.
Australian Prudential Regulation Authority (APRA), 2015. Reporting Standard SRS 7000 Product Dashboard. APRA, Sydney.
Australian Prudential Regulation Authority (APRA), 2017. Annual MySuper Statistics. APRA, Sydney.
Australian Prudential Regulation Authority (APRA), 2018. Quarterly Superannuation Performance. APRA, Sydney.
Australian Securities and Investments Commission (ASIC), 2013. Report 378, Consumer testing of the MySuper product dashboard. ASIC, Sydney.
Australian Securities and Investments Commission (ASIC), 2014. Information Sheet 170: MySuper product dashboard requirements for superannuation trustees. ASIC, Sydney.
Australian Securities and Investments Commission (ASIC), 2017. Report 529, Member experiences of superannuation. ASIC, Sydney.
Ayres, I., Curtis, Q., 2015. Beyond diversification: the pervasive problem of excessive fees and dominated funds in 401 (k) plans. Yale Law J 124, 1476–1552.
Barber, B., Odean, T., Zheng, L., 2005. Out of sight, out of mind: the effects of expenses on mutual fund flows. J Bus 78, 2095–2119.
Bateman, H., Dobrescu, L.I., Newell, B.R., Ortmann, A., Thorp, S., 2016. As easy as pie: how retirement savers use prescribed investment disclosures. J Econ Behav Organ 121, 60–76.
Benartzi, S., 2001. Excessive extrapolation and the allocation of 401 (k) accounts to company stock. J Finance 56, 1747–1764.
Benartzi, S., Thaler, R., 1999. Risk aversion or myopia? Choices in repeated gambles and retirement investments. Manage Sci 45, 364–381.
Benartzi, S., Thaler, R., 2001. Naive diversification strategies in defined contribution saving plans. Am Econ Rev 79–98.
Berk, J., Green, R., 2004. Mutual fund flows and performance in rational markets. J Polit Econ 112, 1269–1295.
Beshears, J., Choi, J., Laibson, D., Madrian, B., 2011. How does simplified disclosure affect individuals' mutual fund choices? In: Wise, D. (Ed.) Explorations in the Economics of Aging. University of Chicago Press, Chicago, pp. 75–96.
Bhandari, G., Deaves, R., 2008. Misinformed and informed asset allocation decisions of self-directed retirement plan members. J Econ Psychol 29 (4), 473–490.
Brown, J., Liang, N., Weisbenner, S., 2007. Individual account investment options and portfolio choice: behavioral lessons from 401 (k) plans. J Public Econ 91, 1992–2013.
Campbell, J., Jackson, H., Madrian, B., Tufano, P., 2011. Consumer financial protection. J Econ Perspect 25, 91–113.
Carlin, B., 2009. Strategic price complexity in retail financial markets. J Financ Econ 91, 278–287.
Carlin, B., Manso, G., 2010. Obfuscation, learning, and the evolution of investor sophistication. Rev Financ Stud 24, 754–785.

Carlin, B., Gervais, S., Manso, G., 2013. Libertarian Paternalism, information production and financial decision making. Rev Financ Stud 26, 2204–2228.
Célérier, C., Vallée, B., 2017. Catering to investors through security design: headline rate and complexity. Q J Econ 132, 1469–1508.
Chant, W., Mohankumar, M., Warren, G., 2014. CIFR Paper No. 020/2014. Centre for International Finance and Regulation, Sydney.
Chetty, R., Friedman, J., Leth-Petersen, S., Nielsen, T., Olsen, T., 2014. Active vs. passive decisions and crowd-out in retirement savings accounts: evidence from Denmark. Q J Econ 129, 1141–1219.
Choi, J., Laibson, D., Madrian, B., 2010. Why does the law of one price fail? An experiment on mutual index funds. Rev Financ Stud 5, 1405–1432.
Colaert, V., 2016. The Regulation of PRIIPs: great ambitions, insurmountable challenges. J Financial Reg 2, 203–224.
Commonwealth of Australia, 2013. Superannuation Legislation Amendment (MySuper Measures) Bill 2013, Explanatory Memorandum.
Consumer Financial Protection Bureau (CFPB), 2012. Know before you owe: evolution of the integrated TILA-RESPA disclosures. Washington, D.C.
Cooper, J., 2010. Super System Review, Final Report – Part Two. Commonwealth of Australia, Canberra.
de Goeij, P., Hogendoorn, T., Campenhout, G., 2014. Pictures are worth a thousand words: graphical information and investment decision making. Mimeo.
Del Guercio, D., Tkac, P., 2002. The determinants of the flow of funds of managed portfolios: mutual funds vs pension funds. J Financial Quant Anal 37, 523–557.
Diacon, S., Hasseldine, J., 2007. Framing effects and risk perception: the effect of prior performance presentation format on investment fund choice. J Econ Psychol 28, 31–52.
Frazzini, A., Lamont, O., 2008. Dumb money: mutual fund flows and the cross-section of stock returns. J Financ Econ 88, 299–322.
Financial Services Council and Association of Superannuation Funds of Australia (FSC and ASFA) (2011) Standard Risk Measure Guidance Paper for Trustees.
Fisch, J., Wilkinson-Ryan, T., 2014. Why do retail investors make costly mistakes: an experiment on mutual fund choice. Univ PA Law Rev 162, 605–647.
Gabaix, X., Laibson, D., 2006. Shrouded attributes, consumer myopia, and information suppression in competitive markets. Q J Econ 121, 505–540.
Gil-Bazo, J., Ruiz-Verdú, P., 2009. The relation between price and performance in the mutual fund industry. J. Finance 64, 2153–2183.
Gillis, T., 2015. Putting disclosure to the test: toward better evidence-based policy. Loyola Consum Law Rev 28, 31–105.
Grinblatt, M., Ikäheimo, S., Keloharju, M., Knüpfer, S., 2015. IQ and mutual fund choice. Manage Sci. https://doi.org/10.1287/mnsc.2015.2166.
Grubb, D., 2015. Behavioral Consumers in Industrial Organization: an Overview. Rev Ind Organ 47, 247–258.
Harries, C., Harvey, N., 2000. Taking advice, using information and knowing what you are doing. Acta Psychol (Amst) 104, 399–416.
Hedesström, T.M., Svedsäter, H., Gärling, T., 2007. Determinants of the use of heuristic choice rules in the Swedish Premium Pension Scheme: an Internet-based survey. J Econ Psychol 28 (1), 113–126.
Henderson, B., Pearson, N., 2011. The dark side of financial innovation: a case study of the pricing of a retail financial product. J Financ Econ 100, 227–247.
Jarvenpaa, S., 1990. Graphic displays in decision making – The visual salience effect. J Behav 3, 247–262.
Jones, L.E., Cäzilia, L., Tennyson, S., 2015. Effects of informational nudges on consumer debt repayment behaviors. J Econ Psychol 51, 16–33.
Kalayci, K., 2015. Price complexity and buyer confusion in markets. J Econ Behav Organ 111, 154–168.
Kalayci, K., 2016. Confusopoly: competition and obfuscation in markets. Exp Econ 19, 299–316.
Kaufmann, C., Weber, M., Haisley, E., 2013. The role of experience sampling and graphical displays on one's investment risk appetite. Manage Sci 59, 323–340.
Kozup, J., Howlett, E., Pagano, M., 2008. The effects of summary information on consumer perceptions of mutual fund characteristics. J Consumer Aff 42, 37–59.
Lipkus, I.M., Samsa, G., Rimer, B.K., 2001. General Performance on a Numeracy Scale among Highly Educated Samples. Med Decis Making 21 (1), 37–44.
Loewenstein, G., Sunstein, C., Golman, R., 2014. Disclosure: psychology changes everything. Annu Rev Econom 6, 391–419.
Lohse, G., 1997. The role of working memory on graphical information processing. Behav Inf Technol 16, 297–308.
Lusardi, A., Mitchell, O.S., 2008. Planning and Financial Literacy: how Do Women Fare? Am Econ Rev 98 (2), 413–417.
Mercer, 2017. Melbourne Mercer Global Pension Index, Report 2017. Australian Centre for Financial Studies and Mercer, Melbourne.
Minifie, J., Cameron, T., Savage, J., 2015. Super Savings. Grattan Institute, Melbourne.
Müller, S., Weber, M., 2010. Financial literacy and mutual fund investments: who buys actively managed funds? Schmalenbach Bus Rev 62, 126–153.
Navarro-Martinez, D., Salisbury, L., Lemon, K., Stewart, N., Matthews, W., Harris, A., 2011. Minimum required payment and supplemental information disclosure effects on consumer debt repayment decisions. J Mark Res 48 (Special Issue 2011), S60–S77.
Pool, V., Sialm, C., Stefanescu, I., 2016. It pays to set the menu: mutual fund investment options in 401(k) plans. J Finance 71, 1779–1812.
Productivity Commission, 2016. Research Report Canberra.
Productivity Commission, 2018. Inquiry Report, No. 19, 21 December 2018 Canberra.
Sirri, E., Tufano, P., 1998. Costly search and mutual fund flows. J Finance 53, 1589–1622.
Vanguard, 2017. How Australia saves: A report on superannuation data. Vanguard Investments Australia.
Vessey, I., 1991. Cognitive fit: a theory-based analysis of the graphs versus tables literature. Decis Sci 22, 219–240.
Walther, T., 2015. Key investor documents and their consequences on investor behavior. J Bus Econ 85, 129–156.
Wilcox, R., 2003. Bargain hunting or star gazing? How consumers choose mutual funds. J Bus 76, 645–665.