NBER WORKING PAPER SERIES

CORRELATION NEGLECT IN STUDENT-TO-SCHOOL MATCHING

Alex Rees-Jones
Ran Shorrer
Chloe J. Tergiman

Working Paper 26734
http://www.nber.org/papers/w26734

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2020, Revised October 2022

For helpful comments, we thank Bnaya Dreyfuss, Rustam Hakimov, Ori Heffetz, Dorothea Kübler, Xiao Lin, Muriel Niederle, Assaf Romm, Alex Teytelboym, and seminar participants at the ASSA Annual Meetings, Microsoft Research NYC, the NBER Summer Institute Workshop on the Economics of Education, New York University, the Niederle Reading Group on Cognitive Reasoning, the Stanford Institute for Theoretical Economics, the Twenty-First ACM Conference on Economics and Computation (EC'20), the University of Toronto. We thank J.P. Bruno and Robert Hovakimyan for excellent research assistance. We thank the Wharton Behavioral Lab and the United States-Israel Binational Science Foundation (BSF grant 2016015) for financial support. The opinions expressed in this paper are solely the authors', and do not necessarily reflect the views of any individual or institution listed above, nor of the National Bureau of Economic Research.

Correlation Neglect in Student-to-School Matching
Alex Rees-Jones, Ran Shorrer, and Chloe J. Tergiman
NBER Working Paper No. 26734
February 2020, Revised October 2022
JEL No. C91,D01,D03,M21

ABSTRACT

We present results from experiments containing incentivized school-choice scenarios. In these scenarios, we vary whether schools' assessments of students are based on a common priority (inducing correlation in admissions decisions) or are based on independent assessments (eliminating correlation in admissions decisions). The quality of students' application strategies declines in the presence of correlated admissions: application strategies become substantially more aggressive and fail to include attractive "safety" options. We provide a battery of tests suggesting that this phenomenon is at least partially driven by correlation neglect, and we discuss implications for the design and deployment of student-to-school matching mechanisms.

Alex Rees-Jones
University of Pennsylvania
The Wharton School
Department of Business Economics
and Public Policy
3733 Spruce Street
Philadelphia, PA 19104-6372
and NBER
alre@wharton.upenn.edu

Ran Shorrer
Department of Economics
The Pennsylvania State University
Kern Building
University Park, PA 16802
rshorrer@gmail.com

Chloe J. Tergiman
The Pennsylvania State University
334 Business Building
University Park, PA 16803
chloejt@gmail.com

A data appendix is available at http://www.nber.org/data-appendix/w26734

A growing body of evidence suggests that many people struggle with decision-making in the presence of correlation. In typical examples of this problem, decision-makers are presented with multiple signals that are each influenced both by independent components and information from a common source. The process by which signals are generated induces correlation, and optimal decision-making requires taking it into account. In practice, however, experiments like those of Enke and Zimmermann (2019) demonstrate that many decision-makers instead act as if these correlated signals were independent—a phenomenon referred to as *correlation neglect*.

In this paper, we study the role of correlation neglect in a decision of considerable importance: the application strategies of students applying to schools. Many application processes inherently require students to make forecasts of events determined by common underlying inputs, resulting in correlation structures like those described above. For example, students commonly must whittle a large number of schools down to a smaller set that are applied to or ranked, introducing an incentive to avoid listing two programs with highly correlated admissions decisions.[1] In such environments, a student harboring correlation neglect faces a difficult decision.

To illustrate this difficulty, consider a simple example. Imagine there are three programs at which you could potentially match, offering payoffs of 3, 2, and 1. Call these programs the best, middle, and worst programs, respectively. These programs will all rank you based on a common, currently-unknown, priority score; assume it will be a random integer drawn from a uniform distribution ranging from 0 to 99. The best program will admit you if your score is at least 50. The middle program will admit you if your score is at least 45. The worst program will admit you with any score. If you could only apply to two of these programs, to which two would you apply?

When considering this problem, one might feel the temptation to apply to the two programs with the highest payouts—we will refer to this as the *aggressive application strategy*. However, doing so is costly in expectation. Because these programs rely on the same score and have near-identical thresholds, the probability of being accepted by the middle program *conditional on being rejected by the best program* is quite low (10%), and insufficient to motivate a risk-neutral decision-maker from taking the sure payoff offered by applying to the worst school. Expected payoff is maximized by applying to the best and the worst programs—we will refer to this as the *diversified application strategy*.

Consider next a slightly modified example. Imagine you are considering the same three programs, but now these programs rank you based on program-specific, statistically independent priority scores. Again, these evaluations are drawn from a uniform distribution ranging from 0 to 99. The best program's acceptance threshold remains at 50, and the worst program continues to admit anyone. However, the middle program's acceptance threshold is changed to 90. In this situation, to which two schools would you apply?

As above, applying to the best and the worst programs is the expected-value-maximizing strategy. Moreover, the consequences of pursuing either the diversified or aggressive application strategies are exactly the same as in the first example. The diversified application strategy grants a 50%

[1] For a recent discussion of optimal diversification strategies in these environments—and their significant complexity—see Shorrer (2019).

chance of enrollment at the best program and a 50% chance of enrollment at the worst program. The aggressive application strategy grants a 50% chance of enrollment at the best program and, conditional on rejection there, a 10% chance of enrollment at the middle program. If one is restricted to these two strategies, choices across these scenarios should be identical.[2]

As we document in this paper, students' application strategies across these scenarios are quite different. When outcomes are correlated, a substantial fraction of students apply to the two most selective programs—i.e., they apply aggressively. By contrast, when priorities are determined independently, subjects intelligently pursue the diversified application strategy at a much higher rate. Despite the numerical equivalency of probabilities, subjects act as if a 10% conditional probability of acceptance (the relevant probability for decision-making in the first example) is much more appealing than a transparent 10% unconditional probability of acceptance (the relevant probability for decision-making in the second example).

These simple examples illustrate something we believe to be a pervasive feature of school choice. In many environments, students can only apply to a subset of the schools that they see as attractive. In such situations, correlation in evaluations at different programs may be neglected or underweighted, leading students to fail to apply the appropriate contingent reasoning when deciding whether to apply to programs of similar selectivity.[3] The consequence is inadequate diversification of application portfolios conceptually similar to the inadequate diversification of asset holdings that is attributed to correlation neglect in Eyster and Weizsäcker (2016).

Concern about decision-quality in the face of correlated admissions is more than academic. As we summarize in Section 1, many countries use school assignment systems that involve choices much like the scenario just considered. In these systems, students are required to submit constrained lists of applications before discovering the results of the standardized test used to determine priority. To the extent that students are unable to correctly reason in such environments, intervention and revision of these systems may be merited.

To study this issue systematically, we conducted our primary study among 165 students of Penn State University. Lab-experimental subjects faced incentivized application scenarios much like the example above. In each scenario, subjects provided a rank-order list (ROL) to be used to match them to one of three schools. These ROLs could only contain two items, however, and thus required the student to choose a school application to forego.

To study the role of correlation neglect, we generated both within-subject and between-subjects variation in the correlation of admissions decisions. The presence or absence of correlation was governed by whether programs' priorities were determined by either a single, common priority score or by program-specific, independent priorities, respectively. Subjects began the experiment by completing nine scenarios across which the acceptance thresholds varied but with the same

[2]Furthermore, while we have emphasized expected-value-maximizing behavior in our example, this equivalence is expected to hold more broadly. Indeed, it should hold so long as preferences depend only on the induced probability distribution over final matches.

[3]For detailed experimental examination of the difficulties of contingent reasoning in the presence of uncertainty, see Martínez-Marquina et al. (2019). Other examples of situations in which subjects fail to properly perform contingent reasoning include Cason and Plott (2014), Esponda and Vespa (2014) and Esponda and Vespa (2019).

correlation condition. They then completed a second battery of the nine scenarios presented under the other correlation condition (with half of subjects facing the opposite ordering). Comparing choices across these conditions provides a clean between-subjects analysis of how application strategies respond to correlated evaluation. A perhaps stronger test arises from the within-subject examination of "matched pairs" of scenarios (like the two example scenarios provided above), in which given strategies result in the same distribution of outcomes in both conditions.

Across these scenarios we document a systematic and quantitatively large tendency to fail to pursue the diversified strategy in the presence of correlation, in effect foregoing desirable "safety" options. We present two classes of evidence suggesting that this relates to incorrect processing of correlated environments. First, to provide a benchmark for correct processing, we presented subjects with direct choices over monetary lotteries. These monetary lotteries were constructed to match the lotteries induced by different application strategies in the scenarios seen by the subject. We find that choices in this transparent domain rationalize the choice of the diversified strategy, and are substantially more predictive of subjects' application strategies when they are made in the absence of correlation. Second, we presented subjects with a variant of the experimental elicitation of correlation neglect of Enke and Zimmermann (2019). We find that this variable predicts subjects' propensity to switch between the diversified and aggressive application strategies in reaction to the correlation of admissions decisions.

To more directly test the role of correlation neglect in this domain, we deploy a second experiment to 165 participants on the Prolific survey platform. This experiment contains the same type of school-choice scenarios as in our primary study, but with two important differences. First, rather than using the "matched pair" design of the primary experiment, we instead vary the correlation condition and the thresholds for admission fully independently. Second, rather than using a comparatively small list of deliberately constructed scenarios, we instead randomly generate the admissions thresholds. The sampling structure for thresholds results in a much larger set of individual scenarios, and is designed in a manner that helps maximize the power of our main analyses.

In these data, we document that the rate of choosing the aggressive strategy only minimally responds to the random assignment of correlation despite the financial incentives for different choices when correlation is present. While a risk-neutral decision-maker would choose the aggressive strategy 22 percentage points more often in the independent arm compared to the correlated arm, the cross-arm difference in our experiment is a mere 3 percentage points.

This experiment is primarily designed to probe the mechanisms underlying this result, and to do so we turn to analyses that are more structural in nature. These analyses involve estimating logit random utility models in which we rationalize ROL choices with the probabilities over school assignments that they induce. In this model, full correlation neglect entails calculating these probabilities as if all priority scores are independent, even when they are not. To test this directly, we calculate the probabilities that would be valid under the priority structures in both the correlated and independent treatment arms, and use both sets of probabilities as predictors in our analysis. We find that choices are well explained by the probabilities calculated assuming independence regardless of whether the choices were made in the independent or the correlated treatment arm. The probabilities calculated assuming correlation have quantitatively much smaller and statistically insignificant effects in both treatment arms. We use this framework to structurally estimate the parameters of a behavioral representative agent model and find that our data are best organized by an estimate of an Enke-Zimmermann-style parameter that suggests full correlation neglect and a Bernoulli utility function with risk aversion. In short, the results of the second experiment directly validate the core prediction of correlation neglect: that choices are made as if subjects calculate probabilities assuming independence even when doing so is not appropriate.

This paper contributes to two literatures. First, and most directly, our paper contributes to the literature on correlation neglect. Common lab-experimental tests of correlation neglect (e.g., Enke and Zimmermann 2019) provide compelling evidence of the underlying behavioral bias. However, in order to isolate the role of correlation and in the interest of being maximally general, these tests are based on abstract forecasting tasks that are several steps removed from most field behaviors of interest. We contribute by identifying a way in which these abstract ideas become concretely relevant for a field behavior of substantial economic importance. We identify a class of large-scale matching systems of interest, provide theory tailored to understanding these environments, and provide tests that directly confirm the application is reasonable. We view this context as a conceptual proving ground for the field relevance of correlation neglect, and view our experimental tests to confirm the need for the integration of these ideas into market design.[4]

Second, this paper contributes to a recently growing literature in "behavioral market design."[5] While work in market design has typically assumed that market participants behave optimally, recent studies from both the lab[6] and the field[7] have shown a meaningful propensity to behave suboptimally. While such studies suggest a role for behavioral economics in the modeling of matching markets, they provide relatively little guidance on the form such models should take. This paper contributes by demonstrating the role of a specific behavioral model capable of making precise in- and out-of-sample predictions about biased respondents' reporting patterns. Such results are necessary to provide theorists with a means of acting on the observation that market participants struggle in these environments.[8] For example, our model directly informs debates surrounding the optimal design of tie-breaking rules. For further discussion, see Section 6.

[4]Note that applications to school choice are not the only suggested field applications of correlation neglect; see also Eyster and Weizsäcker (2016) for applications to financial decision-making and Levy and Razin (2015) for applications to voting behavior.

[5]For a recent reviews of this literature, see Hakimov and Kübler (2020) or Chen et al. (2020).

[6]See Featherstone and Niederle (2016); Guillen and Hakimov (2017) and (2018); Ding and Schotter (2017); Basteck and Mantovani (2018); Li (2017); or Koutout et al. (2021).

[7]See Hassidim et al. (2021); Rees-Jones (2018); Rees-Jones and Skowronek (2018); and Shorrer and Sóvágó (2018).

[8]For other examples of experiments testing the role of specific behavioral models as accounts of mistakes in matching markets, see Li (2017) examining failures of contingent reasoning, Pan (2019) or Dargnies et al. (2019) examining self-confidence, or Dreyfuss et al. (2021) examining expectations-based reference dependence. Note that the models considered in these papers do not predict differences in behavior across our correlated and uncorrelated environment in our primary experiment (conditional on choosing one of the focal strategies constructed to have equivalent payoffs across environments).

The paper proceeds as follows. Section 1 presents summaries of the matching environments that motivate our study and guide our experimental design. Section 2 theoretically formalizes correlation neglect and its consequence of aggressive application strategies. Section 3 presents details of the design and deployment of our primary experiment, and Section 4 presents results of the primary experiment. Section 5 presents the structural experiment. Section 6 concludes by discussing further implications of this bias for market design.

1. Motivating Matching Environments

We begin by describing a set of existing matching systems that help motivate our interest in correlation neglect. While some degree of correlation in admissions decisions is ubiquitous in school-choice environments, we focus on a class of systems where the correlation structure is particularly stark: systems in which application deadlines occur before students learn their performance on standardized tests that determine their priority. To the extent that uncertainty in admission is driven by uncertainty about test performance, this structure results in substantial correlation in admissions outcomes, and ultimately induces a decision problem quite similar to the example considered in the introduction.

Below, we summarize three national school-choice systems with these features, chosen both for their link to our experiment and for the presence of evidence of mistakes in applications strategies.

1.1. **The United Kingdom: The Universities and Colleges Admissions Service.** The vast majority of college admissions in the United Kingdom are organized by the "Universities and Colleges Admissions Service" (UCAS).[9] When participating in the system, aspiring students may apply for up to five courses of study. These applications are due by mid-January, although some courses impose earlier deadlines.[10]

At the time of application, test scores that are used for admissions decisions are not yet available for most of the applicant pool—specifically, A-level exams for those currently finishing their secondary education will typically only be available the following August. In contrast to our simplified example in the introduction, A-level exam scores are not completely random—they may be forecasted based on private knowledge of the student's ability, performance on past exams, and predicted grades that are formally submitted by teachers in this system. Despite such predictive elements, some elements of exam performance remain stochastic, and predicted grades are commonly lamented for their inaccuracy (see, e.g., UCAS 2011, Wyness 2016). Given the imporance of A-level exams and the remaining uncertainty regarding scores, this system is designed to permit educational institutions to make offers of admission *contingent on scoring above a specified*

[9]In 2018, 695,565 applications to undergraduate-degree level courses were received, resulting in 533,360 matches. https://www.ucas.com/data-and-analysis/undergraduate-statistics-and-reports/ucas-undergraduate-end-cycle-reports/2018-end-cycle-report

[10]Oxford, Cambridge, and courses in medicine, dentistry, and veterinary science have application deadlines in mid-October.

threshold when these tests are taken. Based on all information provided in an application package, the institution calculates an individually tailored threshold grade that must be achieved for admission. Nearly all offers take this form.[11]

Due to this structure, students ultimately face a decision of which contingent contracts to pursue, with contracts each offering a subjective probability of admission determined by the students' beliefs that they will surpass their assigned threshold. By the end of March, students will have heard back from all of their five choices. At this time they must specify a "firm" choice and an "insurance" choice and decline all other offers. This is effectively a commitment to attend the firm choice if the conditions of admission are met. If the firm choice's conditions are not met, the student is considered for admission at the insurance choice. If the insurance choice's conditions are then not met, the student is unmatched. While some procedures are in place to assist students who are ultimately unmatched after test scores are observed, students are strongly incentivized to be matched through the straightforward application of this process.[12]

As is readily apparent, students make decisions in this environment facing substantial uncertainty about how they will be evaluated, with this evaluation being correlated across schools. In the first stage, the student must whittle the universe of possible applications one could submit into a list of merely five, understanding that programs will have some degree of similarity in the manner in which they assess the student's extant profile. In the second stage, once offers are received, the student must whittle this set of offers into only two, typically with both of the offers conditioning on a common test score.

Several patterns in application data call the wisdom of students' applications into question. In a review of the system conducted in 2011, the UCAS found that

> "Many highly qualified applicants apply only to a narrow range of very selective [higher education institutions] which find it difficult to differentiate between these applications. This leads each year to a number of candidates with excellent grades failing to gain a place."

This review of the system revealed a number of frictions to perfect use of the system, including a lack of transparency and elements of confusion surrounding the functioning of the system and jargon associated with its implementation. Beyond these types of comprehension problems, however, the largest concerns appear directly related to correlation neglect. Large fractions of program representatives expressed concerns that insurance choices were chosen unwisely in a manner that significant harmed students. Illustrating this worry, 42% of applicants applying before test scores are available list an insurance choice with conditions for admission that are at least as stringent as those for the firm choice, in effect guaranteeing that the student remains unmatched if admission to the firm choice is not secured.[13] This worry led 68% of institutions consulted in the review to

[11] In 2018, only 7.1% of offers were unconditional.

[12] Beyond creating worries about the consequences of correlation neglect, this structure facilitates regression-discontinuity analysis of the consequences of admission (see Broecke 2012).

[13] We say "in effect" because, in some special circumstances, this behavior can be justified; however, these circumstances are extremely rare.

indicate a preference to reform the insurance choice system, with 40% of institutions supporting a policy of preventing students from listing more selective programs in the insurance slot (UCAS 2012).

In summary, this setting contains suggestive evidence of widespread problems consistent with correlation neglect, and actors in this system have been sufficiently concerned with this behavior to seriously consider reforms to mitigate it (see UCAS (2011) for full documentation).

1.2. **Ghana: The Computerized School Selection and Placement System.** In Ghana, applications to senior high school[14] are organized through the Computerized School Selection and Placement System (CSSPS). This system, and problems that arise in students' use of it, is carefully examined in Ajayi (2013); we summarize key elements here.

Since 2005, senior high school admission has been conducted with a deferred acceptance algorithm (Gale and Shapley 1962). Through this system approximately 350,000 students are matched into 700 senior high schools every year. When participating in this match process, students submit rank-order lists of school/program-track pairs. Priorities in the schools are determined by the students' performance on the Basic Education Certificate Exam, which has not yet been taken at the moment of rank-order list submission. After performance on this exam is observed, the algorithm is applied and admissions are announced.

If students were able to list complete preference orderings, the well-known strategy-proofness property of deferred acceptance (Dubins and Freedman 1981, Roth 1982) would absolve students of the need to forecast their admission probabilities at different schools, and thus absolve them of their need to account for correlation in admissions decisions. However, the CSSPS imposes limits on the number of programs a student may rank. Upon initial implementation, students could only rank three choices; this was expanded to four in 2007 and six in 2008. This limit, which binds for the majority of students, introduces strong incentives to mitigate the risk of rejection from all listed programs, and the optimal strategy for choosing a portfolio of ranked schools depends crucially on the correlation structure of admission decisions (Shorrer 2019). As shown in Ajayi (2013), a substantial fraction of students submit rank-order lists with features that are ruled out by optimal behavior. For example, 92% of students ranked schools in an order different than their selectivity, creating situations where rejection by the "back-up" option is assured conditional on rejection by the higher-ranked option. Furthermore, students coming from low-performing schools are more likely to follow these unwise application strategies, suggesting that differences in interactions with the matching mechanism help contribute to the less desirable admissions for students in this group. Ajayi et al. (2020) further demonstrate that the problems associated with these behaviors were only minimally influenced by information interventions, suggesting that these behaviors are not driven by a lack of information but rather by errors in how it is processed.

1.3. **Kenya: Secondary School Admissions.** In Kenya, admissions to secondary school occur through a matching mechanism similar to those described above. At the end of eight years of

[14]Following 6 years of primary school and 3 years of junior high school.

primary-school education, students register and take the national Kenya Certificate of Primary Education (KCPE) examination. As part of the registration process—and crucially, prior to taking the exam—students submit their rank ordering over secondary schools. Government-run secondary schools are grouped into three quality-differentiated tiers: national, provincial, and district schools. Students list up to two choices from each tier, and are admitted to their most preferred option in the highest tier where they may be granted admission. Admissions depends on the outcome of the KCPE test as well as district quotas. For further discussion (on which this summary is based), see Lucas and Mbiti (2012).

As documented in Lucas and Mbiti (2012), patterns similar to those in the UK and Ghana arise. Among the top 5% of students in the 2004 administrative records of the KCPE—i.e., those with a realistic chance of admission to a national-tier school—36% of students listed a second choice school that was more selective than their first choice. As in the UK example, because the second choice is only considered if admission to the first choice is denied, this pattern of reporting effectively foregoes one of only two opportunities for admission at this tier of school. Students making this error reduced their probability of admission to a national-tier school by over two percentage points—a large effect compared to the base admissions rate of 7.2% in the considered sample. While encouraging further study, Lucas and Mbiti conclude that "school choice errors in the admissions process could undermine offering the best opportunities to the highest ability students and cause inequalities to persist."

1.4. **Summary and Interpretation.** We have highlighted three large-scale matching systems with a key feature of interest: requirements to apply to a short list of schools in the presence of substantial uncertainty about a common factor affecting admissions. We note, however, that this structure is not limited to these three domains. Similar matching systems exist in China, Hungary, Trinidad and Tobago, and numerous sub-national applications. In short, this decision-environment is relatively common.

Despite the red flags raised about decision making in these systems, fully assessing the quality of application strategies is challenging. The evidence of mistakes summarized above is limited to cases where a subject lists an option with zero probability of realization—a limited subclass of all mistakes. Focus on mistakes like these is often necessary because the analyst lacks data on the perceived utility associated with different schools. The absence of this data allows one to explain many questionable application strategies with extreme preferences rather than erroneous probability assessments. By contrast, when presenting subjects with experimental scenarios, the subjective value of different schools may be better controlled. When combined with the experimental manipulation of the degree of correlation in admissions decisions, this allows for the precise identification of the errors in reasoning we seek to study.

2. A Theory of School Applications with Correlation Neglect

In this section we formalize our discussion of correlation neglect. We state precisely its meaning in the context of school choice, then we establish its consequences for subjective expected utility and for preference submission. In the interest of proceeding to our experimental results quickly, we present our propositions with only brief intuitive explanation and relegate all formal proofs to Appendix A.

2.1. Model Preliminaries.

Consider a set X of schools. For simplicity, we assume that, conditional on the information available to students at the time of application, schools' admissions decisions are based on a single exam. Each agent has beliefs about how he will perform on the exam, summarized by the CDF F,[15] and he knows, for any $s \in X$, his utility from attending this school, u_s, and the score threshold required for admission, c_s. The utility from being unassigned is normalized to zero. Taken together, the set of available schools and the vector of corresponding admission thresholds form a *school-choice environment*, formally denoted by $E = (X, c)$.

In the leading example in the introduction, $X = \{\text{best, middle, worst}\}$; $c_{\text{best}} = 50$, $c_{\text{middle}} = 45$, and $c_{\text{worst}} = 0$; and $u_{\text{best}} = 3$, $u_{\text{middle}} = 2$, and $u_{\text{worst}} = 1$. In this example, the agent believes that scores are distributed uniformly over the integers between 0 and 99.

A *rank-order list* (ROL) is an ordered list of schools. Upon submission of an ROL, a student is admitted to the highest-ranked school at which admission is granted. Such ROLs are formally used in centralized matching markets applying, e.g., the deferred acceptance algorithm or its variants. Furthermore, Shorrer (2019) observes that ROLs may be considered to implicitly exist in decentralized school-choice markets, in which students attend the best school that accepts them.[16] Given an ROL, r, and an integer, i, we denote by r^i the i-th ranked school on that ROL. If the ROL r ranks school j higher than school k, we denote this relationship by $s_j \succ_r s_k$. We say that an ROL is *undominated* if schools included in the ROL are ordered according to true preferences (i.e., $u_{r^i} \geq u_{r^{i+1}}$ for all $i < |r|$).[17]

2.2. Correlation Neglect, Expected Utility of ROLs, and Chosen Application Strategies.

In this decision problem, we assume that students evaluate the value of an ROL using standard subjective expected utility. Formally, this subjective expected utility is governed by the equation

$$\sum_{s \in r} Pr(\text{rejection at all } s_i \succ_r s) \cdot Pr(\text{admission at } s | \text{ rejection at all } s_i \succ_r s) \cdot u_s. \tag{1}$$

[15] Unless otherwise mentioned, we assume, without loss of generality, that students beliefs about scores are uniform on the unit interval.

[16] This holds since students will only attend lower-ranked (less desirable) schools if they are rejected by all higher-ranked (more desirable) schools. Consequently, optimal ROLs can be calculated using similar dynamic programming as applies to the centralized case, and lower-ranked schools should be chosen conditional on the student being rejected by all higher ranked schools.

[17] Note that, for any set of schools one applies to, the ROL that orders them according to true preferences yields the highest expected utility (under the definition presented in the following section). This holds regardless of the agent's risk preferences and beliefs (including about correlation).

We consider two types of agents, differing in their assessment of subjective expected utility. As a benchmark, we consider *sophisticated* agents. By assumption, these agents correctly evaluate all probabilities in Equation 1. We contrast this type against *fully correlation neglectful* agents (sometimes referred to as neglectful agents in shorthand), who treat admissions to each school as if they were independent. This assumption leads students to replace the term $Pr(\text{admission at } s| \text{ rejection at all } s_i \succ_r s)$ with the term $Pr(\text{admission at } s)$, and to compute $Pr(\text{rejection at all } s_i \succ_r s)$ differently in situations with two or more prior rejections occurring. When we refer to the *subjective expected utility* of the neglectful type, which we denote by $V_n(r)$, we refer his expected utility based on his misguided probability assessments. When we refer to the *experienced utility* of either type (as well as the subjective expected utility of the sophisticated type), we refer to correctly evaluated expected utility, denoted $V_s(r)$.

Given these definitions, our use of the term "correlation neglect" may best be understood not as a reference to a fundamental, underlying bias, but as a reference to a reduced-form phenomenon: cases where correlation in outcomes necessitates Bayesian updating and the relevant updating is not pursued. It is worth noting that several different underlying mistakes in reasoning could generate this behavior. For example, an agent who completely understands the relevant correlation structure may fail to see any need for contingent reasoning. Alternatively, an agent who completely understands contingent reasoning may fail to see that correlation exists. While both examples result in neglecting the consequences of correlation, they derive from quite different underlying misunderstandings. While prior research has worked to disentangle the specific errors underlying these probabilistic judgments (see, e.g., Levin et al. 2016), in our environment the distinction does not result in differing predictions.

Illustrating these definitions in the context of our leading example, consider the ROL (best, middle)—the aggressive application strategy. The neglectful agent's subjective expected utility from this ROL is $V_n\,(\text{best, middle}) = .5 \times 3 + (1 - .5) \times .55 \times 2 = 2.05$. His expected experienced utility, which is equal to the sophisticated type's expected utility, is $V_s\,(\text{best, middle}) = .5 \times 3 + (1 - .5) \times .1 \times 2 = 1.6$. Because $V_n\,(\text{best, middle}) \geq V_s\,(\text{best, middle})$, the neglectful agent perceives a higher expected utility than he would if he were sophisticated. This relative optimism is not a coincidence, as we illustrate in Proposition 1.

Proposition 1. *For any school-choice environment and for any undominated rank-order list r, $V_s(r) \leq V_n(r)$.*

Put simply, because the neglectful agent fails to account for the "bad news" that a rejection conveys about the as-yet-unknown test score, he overestimates his chances of admissions after such a rejection occurs. This results in an overestimation of expected utility.

We now fully define behavior that we wish to characterize and study: ROL choice that maximizes the notion of subjective expected utility just established. Let $r(k, u, E)$ denote the optimal size-k ROL for agent with preferences u in environment E.[18] When E and u are clear, we often just write $r(k)$. Similarly, $r_n(k,u,E)$ and $r_n(k)$ denote the perceived optimal ROL of the neglectful agent.

To illustrate in our leading example, we calculate the subjective expected utility associated with the three undominated ROLs:

$$\begin{aligned} V_n\,(\text{best, middle}) &= .5\times 3+(1-.5)\times .55\times 2 &&= 2.05\\ V_n\,(\text{best, worst}) &= .5\times 3+(1-.5)\times 1\times 1 &&= 2\\ V_n\,(\text{middle, worst}) &= .55\times 2+(1-.55)\times 1\times 1 &&= 1.55 \end{aligned}$$

Thus, the neglectful agent will chose (best, middle) over (best, worst) and (middle, worst). Formally, $r_n(2) = (\text{best, middle})$. However, experienced utility is given by:

$$\begin{aligned} V_s\,(\text{best, middle}) &= .5\times 3+(1-.5)\times .1\times 2 &&= 1.6\\ V_s\,(\text{best, worst}) &= .5\times 3+(1-.5)\times 1\times 1 &&= 2\\ V_s\,(\text{middle, worst}) &= .55\times 2+(1-.55)\times 1\times 1 &&= 1.55 \end{aligned}$$

.

Choosing an ROL to maximize V_n thus guides the agent to choose the aggressive application strategy (best, middle) when the diversified application strategy (best, worst) is objectively utility maximizing. The agent is expected to lose .4 experienced utils due to this mistake.

We next demonstrate that the consequences of correlation neglect for experienced utility may be grave. To do so, we first define a notion of the price of neglect that captures the fraction of experienced utility lost by the neglectful type.

Definition 1. *The price of neglect for a neglectful agent with utility u in environment E subject to constraint k is equal to the the difference in experienced utility between the maximizing size-k ROL and the ROL chosen by the neglectful type, normalized by the expected experienced utility from the maximizing ROL. In formal notation,* $PN(u,E,k) = \frac{V_s(r(k)) - V_s(r_n(k))}{V_s(r(k))}$.

Applying this definition, in the worst case, an optimal ROL of size k may generate k-times more experienced utility than the one that maximizes the subjective expected utility.

Proposition 2. *For any integer k, and any decision environment where the agent is constrained to (costlessly) apply to up-to-k schools, the price of neglect for the neglectful type is bounded above by $1-\frac{1}{k}$. Furthermore, this bound is tight—for any k, there exist school-choice environments where the price of neglect is arbitrarily close to $1-\frac{1}{k}$.*

[18]For simplicity, we assume that both the sophisticated and the neglectful type have a unique optimal ROL. This assumption, which is satisfied generically, plays no role in the analysis, and is only made to simplify statements.

To illustrate how this worst-case bound may be achieved, consider a modification to our leading example. In this modification we add one more school to the choice set, and this school yields the same utility and has the same admissions threshold as the best school. The neglectful agent would treat this copy of the best school as another (independent) chance for admission, ignoring the fact that rejection by one copy guarantees rejection by the other. As a result, the second application on his ROL is wasted, and he is no better off than he would be applying to a single school. As we show in Appendix A, for any permitted length of ROLs (k), we can construct examples involving perfect substitutes in which the neglectful agent will apply in a way that makes him no better off than if he had an ROL of length 1. Furthermore, a sophisticated agent can achieve approximately k times higher utility because the optimal length-k ROL in the examples we construct achieves approximately k times the utility of the optimal length-1 ROL.

While these extreme examples rely on the existence of perfect substitutes, note that the common situation of imperfect substitutes generates a similar conceptual force. In our leading example, because the best and middle programs have very similar thresholds for admission, a rational agent should be hesitant to apply to both, and the failure to see this reasoning drives the utility losses documented above.

Given these observations, we conclude with a final result formally establishing the sense in which neglectful application strategies are overly aggressive.

Proposition 3. *For any constraint on the size of the ROL k, the neglectful type is at least as likely to be unassigned as the sophisticated type.*

In our leading example, recall that the sophisticated type would submit the ROL (best, worst), whereas the neglectful type would submit the ROL (best, middle). Because the worst school guarantees admission in the example, the sophisticated type faces no risk of being unassigned. By contrast, the neglectful type faces a 45% chance of remaining unassigned. In our example, this is characterized as an aggressive strategy: it contains options with higher utility, conditional on assignment, at the cost of exposing the agent to a greater degree of downside-risk of remaining unassigned. Proposition 3 demonstrates that the pursuit of more aggressive strategies is not unique to the example, but rather a general feature of ROL choice among neglectful agents.

2.3. **Summary.** The presence of correlation neglect leads subjects to be overly optimistic about their chances of admission to schools that they rank below their first choice. As a result, they undervalue the need to diversify the portfolio of schools contained in their ROL, resulting in overly aggressive application strategies. We experimentally test these predictions in the sections that follow.

3. Design and Deployment of the Primary Experiment

3.1. **Summary of Experimental Materials.** In this section, we summarize all measures and manipulations included in our primary experiment. All experimental materials are available in the Materials Appendix.

The experiment began with a brief informed consent document. Subjects were then told that the experiment was divided into parts (which we will refer to as modules), and that decisions in any part would not affect the opportunities presented in any other. Throughout the experiment, paper instructions were distributed and read out loud by the experimenter and subjects were given the opportunity to ask clarifying questions. The relevant experimental elicitations were then presented through a Qualtrics interface.

3.1.1. *Incentivized School-Choice Scenarios.* After reviewing the introductory materials, subjects were presented with the school-choice scenarios of primary interest. Within each scenario, students faced three programs to which they could apply. We referred to these programs as Colleges A, B, and C, and subjects could match to no more than one of them. To dictate the desirability of matching to these programs, each yielded a different payoff to matriculating students. Students matriculating to A, B, and C would receive a bonus payment of \$10, \$5, and \$2.5, respectively. Assignment to programs was determined by a matching procedure that depended on both test scores and students' rank-order lists (ROLs).

Test scores were simulated with draws from a uniform distribution ranging from 0 to 99, a structure known to participants. In the correlated-admissions module, a single test score was used for all programs' admissions decisions. In the uncorrelated-admissions module, three statistically independent tests governed admissions to the three programs. Minimum test score requirements were presented alongside each school's bonus payments and are varied across scenarios.

Based on this information, students were faced with the task of choosing an ROL to be used in the assignment procedure. ROLs were ordered lists of two of the three schools; building an ROL required choosing one school application to forego and choosing an ordering among the remaining two applications that are submitted. The student was paired to the highest-ranked program for which the admission threshold was passed. In all cases, test scores were realized after ROLs are determined.

Construction of Scenarios: Our scenarios were constructed to function as "matched pairs," under which equivalent payoff structures were induced either in a correlated or uncorrelated decision environment. Table 1 summarizes each matched pair of scenarios, presents lotteries induced by the two focal ROLs, and reports the expected value of those lotteries. To illustrate, consider the first scenario. When outcomes were determined by a single priority (i.e., in the correlated-admissions module), the first scenario had a score threshold of 50 for school A, 45 for school B, and 0 for school C. When outcomes were determined by multiple, independent priorities (i.e., in the uncorrelated-admissions module), the thresholds were 50, 90, and 0. This example matches the scenario presented in the opening paragraphs of the paper. The aggressive application strategy ($(A \succ B)$) and the diversified application strategy ($(A \succ C)$) result in the same probabilities of admissions at each school and thus equivalent expected payouts, summarized in right columns of the table.

We constructed our nine matched pairs with several considerations in mind. While we were initially motivated by pilot results arising from scenario 1, we wanted to ensure that the patterns of behavior we observed were not somehow unique to the thresholds in that scenario.

First, we were concerned that some of the applications to program C might be due to the attraction of a completely certain option. This motivated the creation of scenario 2, which closely mirrors scenario 1 but makes admissions to the worst program uncertain (but still very likely). As we vary other thresholds across additional scenarios, we continue to create pairs that differ only in the certainty of admission to the third program (see scenario pairs (3, 4), (6, 7) and (8, 9)).

In scenarios 1 and 2 (as well as all other scenarios we will discuss), pursuing the aggressive application strategy ($(A \succ B)$) induces a lottery that is both riskier and (weakly) lower in expected value than the diversified application strategy ($(A \succ C)$). In scenarios 3 and 4, we set the score thresholds in order for the aggressive application strategy to yield a higher expected value, making it potentially desirable for some risk preferences.

We constructed scenario 5 to study the extreme type of mistakes observed in the field settings described in Section 1: submitting second-choice options for which rejection is guaranteed conditional on rejection by the first choice. In the correlated-admissions module, the required test score for the middle program was 55, whereas the required score for the best program was 50. In the uncorrelated-admissions module, the independent priority score necessary for admission to the second program was 100. In both modules, applying to the top two programs yielded a 50% chance of admission to the best program and a 0% chance of admission to the middle program. Submission of the preference order $(A \succ B)$ therefore mirrors the worrying behaviors seen in the UK, Ghana, and Kenya.

Note that all of scenarios 1-5 are designed to focus on the pursuit of ROLs $(A \succ B)$ and $(A \succ C)$: the remaining undominated preference order ($(B \succ C)$) is not meant to be appealing and empirically is rarely chosen. Scenarios 6-9 were included to examine application behavior in cases where the ROL $(B \succ C)$ is made more attractive (although our focus remains on preferences $(A \succ B)$ and $(A \succ C)$).

Examining responses to the variations across these scenarios provides a means of testing whether ROL submissions respond to relatively simple incentives, which can assuage some concerns that the behaviors we document are driven by a lack of understanding or responsivity to scenario details. Additionally, and more importantly, across these scenarios we may examine how the pursuit of the aggressive and diversified application strategies responds to correlation when a battery of other considerations are varied. This helps ensure that any response that we observe is not unique to a particular arrangement of parameters.

These scenarios were divided into two blocks of nine, with the correlation structure constant within each block. Which correlation structure subjects saw first was randomized at the session level. The order of questions within block was randomized at the subject level.

3.1.2. *Auxiliary Measures and Questions.* Following the school-choice scenarios, three additional groups of questions were presented.

Preferences over Lotteries: First, subjects were presented with a series of nine choices over risky lotteries. These lotteries were constructed to match the lotteries over monetary outcomes induced by the two focal admissions strategies ($(A \succ B)$ and $(A \succ C)$) submitted in each of the nine school-choice scenarios, as seen in Table 1. By eliciting direct preferences over these lotteries, we may benchmark the choices made in the school-choice scenarios against choices that are made when their consequences are fully transparent.

Raven's Matrices: Second, subjects were presented with a battery of "Raven's Progressive Matrices," a common assessment of spatial reasoning used as a proxy for general cognitive ability (Raven and Raven 2003). Subjects were given five minutes to complete as many of the 6 matrices as they could.

Direct Elicitation of Correlation Neglect: Third, subjects faced a correlation-neglect elicitation closely derived from that used by Enke and Zimmermann (2019). Subjects were given the task of forecasting an underlying value, denoted "X." X was drawn from a normal distribution with a mean of zero and a standard deviation of 500. Subjects were asked to guess the value of X and were compensated based on their accuracy. The probability of winning a $10 bonus was governed by the squared difference between the subjects' estimate and the true value, providing incentives for truthful reporting. To help guide this decision, four noisy signals of the true value were drawn, and were communicated to the subject by "communication machines" (CMs). One of the four signals was observed by all four CMs, and their reporting patterns induced correlation into their communicated forecast. One CM directly reported the common signal, whereas the other three reported the average of the common signal and a signal only observed by that CM. All details of the noise distributions, signal generation, and reporting structure were communicated to participants.

As shown by Enke and Zimmermann (2019), this environment offers a clear way to measure the degree of correlation neglect. The optimal forecast in this environment is constructed by using the known correlation structure to back out the 4 signals provided to the CMs, then average those signals. Denote this optimal forecast as f^o. Alternatively, one could imagine a subject treating the four reports of the CMs as if they were four independent signals and simply averaging them. Denote this naive forecast as f^n. As long as $f^n \neq f^o$, any individual forecast maps onto a specific value of χ implicitly defined by the equation $f = \chi f^n + (1-\chi) f^o$. Enke and Zimmermann use χ as a measure of the degree of correlation neglect in subjects' forecast, noting that $\chi = 0$ corresponds to a completely optimal forecast and $\chi = 1$ corresponds to treating the data as if it were independent.

To help mitigate measurement error, we follow Enke and Zimmermann's strategy of offering subjects multiple forecasting tasks and assigning them the median of their measured χ values. Subjects completed 3 forecasts.

3.1.3. *End of experiment.* The experiment concluded with a brief elicitation of demographics. Following these questions, bonuses for incentivized modules were determined and final payments for the experiment were made.

3.1.4. *Compensation.* Subjects received a show-up fee of $7.00. In addition, subjects were incentivized to truthfully report their preferences and carefully answer questions. The incentives for each module were explained prior to its presentation, and are summarized here:

- One round was randomly chosen among the school-choice scenarios and the equivalent "preferences over lotteries" questions. If that randomly chosen round was one of the school-choice scenarios, the student's ROL was processed. Earnings consisted of the bonus associated with the school that a student was admitted to. If the randomly chosen round was one of the lottery questions, we ran the lottery that the subject selected and earnings consisted of the outcome of the selected lottery.
- Subjects also received $1.00 for each correctly solved Raven's Matrix.
- One of the three direct correlation neglect elicitation questions was randomly selected. A subject received an additional $10 if the submitted forecast was "close enough" to the true value.[19]

Subjects were informed of their earnings in each module only at the end of the session.

Average total earnings were $18.10, and ranged from $7.00 to $33.00 (including the $7.00 show-up fee).

3.2. **Preregistration.** We preregistered our hypothesis that correlation would increase the propensity of the aggressive strategy, our primary analyses, our target sample size, and our sample inclusion rules prior to the beginning of data collection. Our preregistration is archived on aspredicted.org, and is included in the Materials Appendix for ease of reference. As we report results, we flag any and all places where we deviate from our preregistered analysis plan.

3.3. **Deployment of the Primary Experiment.** Our experiment was conducted in January and February 2019 at the Laboratory for Economics, Management and Auctions (LEMA) at Penn State University. Experimental sessions took approximately one hour, and subjects participated in one session only. We recruited 80 subjects in the "correlated admissions first" treatment and 85 in the "uncorrelated admissions first" treatment, consistent with our preregistered target of 80 per cell. The treatments differed only in whether students saw the correlated or uncorrelated questions first.[20]

Basic demographics of our sample are presented in Appendix Table A1.

[19] As in Enke and Zimmermann (2019), the threshold for "close enough" was determined by a random draw.

[20] Because we generally do not detect order effects (under which the distribution of, e.g., correlated choice would depend on whether the module appeared first or second), we present analyses which pool questions of the same type regardless of position in the experiment. In Appendix B we present analyses supporting this decision. This analysis consists of formal tests for order effects (Appendix Table A2) and a recreation of our main analysis using only data from the first module seen (Appendix Table A3). The latter analysis additionally serves as a between-subjects replication of the primary results we discuss in text.

4. Results of the Primary Experiment

4.1. **Application Strategies in Scenarios.** Table 2 summarizes the application strategies pursued in each of our scenarios. Examining the first row, we see that when the scenario considered in the introduction is presented with correlated admissions decisions, 44.9% of subjects pursued the diversified application strategy ($(A \succ C)$). Among students submitting another ROL, by far the most common was $(A \succ B)$—the aggressive application strategy. 48.5% of students pursued this strategy despite its greater risk and lower expected value.[21] This behavior—which we interpret as a mistake—is substantially less prevalent in the uncorrelated-admissions module. In this environment, only 10.9% of students submitted ($A \succ B$), with 84.2% of students making the optimal choice of ($A \succ C$). In short, in this scenario, subjects were more tempted to pursue the aggressive (and perhaps unwise) admissions strategy of ($A \succ B$) in the correlated admissions environment.

To assist in assessing these claims statistically, the final three columns of the table present p-values arising from a set of cross-module hypotheses tests. The first column presents a Fisher's exact test of whether the distribution of chosen ROLs varies by correlation module. The second and third columns present two-sample difference-of-proportions tests of equality in the fraction choosing the aggressive and diversified application strategies. Examining these statistics for the first scenario demonstrates that all differences discussed in the prior paragraph are unlikely to arise by chance.[22]

Across the different thresholds induced across the nine scenarios, the patterns described above always holds: the strategy of diversifying to the best and the worst programs is more likely to be pursued in the uncorrelated-admissions module, and our target "tempting" behavior of aggressively applying to the top two programs is more likely in the presence of correlated evaluations. This remains true when the worst program still has residual uncertainty of admission (as in scenarios 2, 4, 7, and 9); when reducing risk by applying to the worst program comes at some cost in expected value (as in scenarios 3 and 4); when admission to the middle program is impossible conditional on rejection by the best program (as in scenario 5), and in environments in which the ROL $(B \succ C)$ is made more attractive (as in scenarios 6-9).

Across these scenarios, we see a variety of patterns indicating responsivity to the incentives introduced by these variants. More subjects submit preference order ($A \succ B$) when it yields a higher expected value than the alternative (as in scenarios 3 and 4), fewer subjects submit ($A \succ B$) when it is a dominated strategy (as in scenario 5), and more subjects submit $(B \succ C)$ when it is made attractive (as in scenarios 6-9). And yet, across all these variants and despite these signs of intelligent response to incentives, substantially different patterns of reporting are seen based on the presence or absence of correlation in evaluation.

[21]Formally, the distribution of outcomes resulting from the ROL ($A \succ B$) is second-order stochastically dominated by that resulting from ($A \succ C$), meaning it should not be chosen by any risk-averse expected-utility maximizer.

[22]In addition to the tests considered in this table, we also conducted signed rank tests. For all comparisons, these are also statistically significant at the 1% level.

Figure 1 summarizes these differences by presenting the distribution of chosen ROLs averaged over all nine scenarios. On average, the rate of pursuit of the aggressive application strategy increases by 20.1 percentage points when scenarios are presented with correlated admissions decisions. This difference is relatively stable across the course of the experiment, with average rates of choosing the aggressive and diversified strategies only minimally varying by presentation order (see Appendix C for documentation).

Patterns like these are additionally seen in within-subject evaluations. For each question, Table 3 shows the pairs of strategies submitted by each subject, focusing specifically on the two focal strategies of $(A \succ B)$ and $(A \succ C)$. The first two columns present the fraction of subjects choosing a focal strategy and not responding to the presence of correlation: these subjects pursue the same admissions strategy in both of the matched decisions. The next two columns present subjects who switched between focal strategies depending on the correlation condition. As illustrated in these columns, substantial fractions of subjects had their answer vary depending on the presence of correlation, with a strong tendency to chose the aggressive strategy $(A \succ B)$ in the correlated-admissions module and the diversified strategy $(A \succ C)$ in the uncorrelated-admissions module. Subjects switched from the aggressive strategy in the correlated-admissions module to the diversified strategy in the uncorrelated-admissions module in 21.1% of subject-scenario pairs, indicating a substantial propensity of this specific preference reversal. As is indicated in the final column, across most scenarios the majority of subjects who submitted the ROL $(A \succ B)$ under correlation would submit the ROL $(A \succ C)$ under independence. In contrast, the mirroring inconsistency (choosing $(A \succ C)$ under correlation and $(A \succ B)$ under independence) never exceeds 6.0%.

4.2. **Lottery Choices.** As presented in Chade and Smith (2006), optimal behavior in this class of decision problems involves recursive consideration of probabilistic outcomes. If subjects fail to appropriately consider and calculate these probabilities, the preferences that they express through their portfolio choices will differ from the preferences they would express when the relevant probabilities are transparently presented for them. To test for these discrepancies, subjects in our experiment faced a series of questions in which they directly chose between the pairs of lotteries presented in the right-most columns of Table 1. These decisions directly elicit preferences over the transparent monetary consequences of submitting $(A \succ B)$ or $(A \succ C)$.

As shown in Figure 2, application strategies were substantially less aligned with lottery choices in the presence of correlation. In the correlated-admissions module, application decisions were consistent with the chosen transparent lottery in 44.0% of cases. In the uncorrelated-admissions module the rate of consistency between application strategies and chosen lotteries rises to 66.9%, a nearly 23 percentage point increase.

Table 4 presents several additional statistics characterizing subjects' preferences over lotteries. As illustrated in the first column, when payoffs induced by the aggressive and diversified strategy were made transparent, subjects overwhelmingly chose in favor of the latter. This supports our interpretation of the pursuit of that ROL as optimal and the pursuit of other ROLs as misguided. In the second column, we use this measure to help resolve which ROL should be honored among the subjects whose strategy differed across modules. Among subjects who chose $(A \succ B)$ in the correlated-admissions module and $(A \succ C)$ in the uncorrelated-admissions module—i.e., the individuals of interest in the previous subsection—we find that relatively few chose the lottery associated with $(A \succ B)$. Only in scenario 4—one of the two in which the expected value of $(A \succ B)$ exceeds that of $(A \succ C)$—do preferences on average tilt towards the aggressive strategy. In the third column in the table, we look at the $(A \succ C)$ choices in the lottery questions for those subjects who consistently chose $(A \succ C)$ in the school-choice setting, providing validation from a third source of the consistency in their responses.

In summary, subjects' preferences over ROLs in the uncorrelated-admissions module tend to align with their preferences inferred from transparent presentations of the induced lotteries. This corroborates our treatment of the more aggressive application strategies in the correlated decision environment as arising from a mistake. Furthermore, because risk preferences inferred from transparent lotteries demonstrates a greater degree of risk aversion than preferences over ROLs, these results suggest a potentially worrying factor influencing the estimation of structural models in school-choice settings.

4.3. **Predicting the Pursuit of Aggressive Strategies with the Enke and Zimmermann Correlation Neglect Measure.** To further validate the relationship between our behavior of interest and correlation neglect, we asked subjects 3 questions taken directly from the materials of Enke and Zimmermann (henceforth EZ; 2019). Following their technique, we compute their parameter of correlation neglect for each of those questions and assign to each subject their median value. Because their parameter is meant to be interpreted on the unit interval—with a value of zero implying completely correct processing and a value of 1 implying that correlated signals are treated as perfectly independent[23]—we restrict attention to cases where this measure falls in this range.[24]

Our goal in these analyses is to use the EZ measure to predict the propensity of the within-subject preference reversals that we argue are generated by correlation neglect. To that end, we

[23]To further explain with a concrete example: consider a case where the optimal forecast (accounting for correlation) were 5, but the simple average (which would provide the optimal forecast if all signals were independent) were 6. The EZ measure would assign the value $\chi = 0$ for a subject who reported 5 and $\chi = 1$ for a subject who reported 6. A subject who reported 5.5 would be interpreted as being partially influenced by correlation neglect and assigned the value $\chi = 0.5$. In this example, measured values of χ outside of the unit interval imply that the subject made a forecast either below 5 or above 6. Such forecasts are indeed evidence of some mistake in forming a forecast, but such mistakes are not naturally attributable to a tendency to treat observations as independent.

[24]We note that we did not preregister this treatment of outlier values of the EZ measure because we did not anticipate their prevalence or importance. In our data, 9% of subjects have an EZ measure below 0, ranging to a minimum value of -0.845. More worryingly, 31% of subjects have an EZ measures exceeding 1, ranging to a maximum value of 1.837. Having a non-unit-interval EZ measure is predictive of our preference reversal of interest, regardless of whether the measure is below zero or above one. We believe that this is because invalid EZ measurements are an indicator of particularly inattentive or confused participants, who would naturally be predicted to make misguided choices at a high rate. Adding observations with a high rate of preference reversals for both the lowest and highest EZ values would naturally obscure any association, and indeed our regression results are weakened when these extreme EZ values are not excluded. We further note that this relatively high frequency of measurements outside of the unit interval is not unique to our study: Figure 2 in EZ demonstrates that they too regularly estimated values outside of the unit interval, also with a greater propensity to estimate values above 1 than below 0.

construct a measure that compares the rate of optimal decisions versus the rate of the specific within-subject mistake that we considered in the prior section. Other ROL reporting patterns are effectively disregarded.[25]

Figure 3 presents a local-polynomial regression of the relative rate of our preference reversals of interest on the Enke-Zimmermann correlation neglect measure. As illustrated in this figure, increasing from zero EZ correlation neglect to full EZ correlation neglect is associated with a substantial rise in the rate of choosing the aggressive strategy under correlation and the diversified strategy under independence (relative to the rate of consistently choosing the optimal strategy across correlation conditions). This association is large and statistically significant in a simple OLS regression, and remains so when we control for our Raven's Matrices measure of cognitive ability and the full battery of demographic variables collected in the study (the results are reported in Appendix Table A4). Beyond the EZ measure, the only additional variable found to be predictive is cognitive ability (a variable commonly found to be associated with making mistakes in matching mechanisms, see Basteck and Mantovani (2018); Hassidim et al. (2021); Rees-Jones (2018); Rees-Jones and Skowronek (2018)).[26]

In summary, the key behavior we have attributed to correlation neglect—pursuing the aggressive application strategy when admissions decisions are correlated and the diversified strategy when they are not—is predicted by existing experimental measures of correlation neglect.

4.4. **Summary of Additional Analyses and Discussions.** In Appendix D we provide further discussion of the robustness of our inferences to alternative models. We document that our results cannot be well explained by aversion to schools dominated as singleton applications (D.1), common models of choice-set dependence (D.2), independence neglect (D.3), or preferences for randomization (D.4).

In Appendix E, we present results that inform whether the tendencies driven by correlation neglect can be mitigated with debiasing interventions. To assess this question, we ran a 1,999 subject online debiasing experiment presenting school-choice decisions of the same structure as those in our primary experiment. In correlated and uncorrelated control conditions, we replicate our finding of more aggressive and less diversified applications in the presence of correlation. Across 5 treatments arms, we presented subjects facing correlation with different debiasing interventions aimed to help improve decision making. While our interventions generally lead to behaviors closer to the

[25]Formally, our dependent variable takes the value of 1 for a matched pair when the subject chose ($A \succ B$) in the correlated-admissions module and ($A \succ C$) in the uncorrelated-admissions module—the behavior that we attribute to correlation neglect—and zero when they chose ($A \succ C$) in both modules—the optimal behavior. All other paths are ignored. Within-subject, we then calculate the fraction of cases in which the subject pursued the correlation-neglectful path. The mean (median) number of observations per subject used to calculate this fraction is 4.9 (5). Note that this measure is undefined for the 8 subjects who followed neither the optimal nor the correlation neglectful reporting pattern for any of the scenarios considered. Their exclusion leads to a usable sample of 157 observations for these analyses.

[26]We additionally preregistered that we would examine the relationship between our within-subject-mistakes dependent variable and Raven's task performance without additional controls. In this analysis we find a similar negative relationship ($\beta = -0.059$, s.e.$= 0.024$, $p = 0.015$).

uncorrelated baseline (our benchmark for behavior not influenced by correlation neglect), the improvements are mostly quantitatively small and statistically insignificant. An important exception is a debiasing intervention modeled after that used by the UK's Universities and Colleges Admissions Service (UCAS), which we find reduces the effect of correlation by approximately 50%. Overall, the results of this auxiliary experiment demonstrate that interventions aimed to help students overcome correlation neglect can have some positive impact. At the same time, these results demonstrate that the patterns we demonstrate are relatively resistant to debiasing: the problem is not simply resolved with slightly different phrasing or additional explanation, and quantitatively large impacts of the bias remain even after the current best-practices in explanations are pursued.

5. Structural Experiment

The results of our primary experiment provide clear evidence of mistakes in reasoning. These mistakes are driven by apparently overaggressive application strategies in the presence of correlation—a predicted consequence of correlation neglect in the theory in Section 2—and are associated with the Enke-Zimmermann measure of correlation neglect—a predicted result if both our mistakes of interests and the mistakes measured by Enke and Zimmermann have the same underlying cause. While these results suggest correlation neglect as an explanation for our findings, they do so through indirect means. In this section, we report the results of an additional online experiment designed to provide direct evidence of the underlying mechanism that we have proposed: reliance on probabilities calculated assuming independence even in correlated environments.

5.1. **Design and Deployment of Structural Experiment.**

5.1.1. *Summary of Experimental Materials.* The structural experiment was built around the same style of incentivized school-choice scenarios used in our primary experiment, and we sought to keep the instructions and presentation of the scenarios closely comparable. The high-level difference in the design involves the specific score thresholds that are presented in each individual scenario. In the primary experiment, our "matched pair" design entails presenting a small set of carefully constructed scenarios, with these scenarios featuring dependency between the score thresholds and treatment assignment. In the structural experiment, we instead present scenarios with score thresholds that are randomly generated statistically independently from treatment status. Our procedure for randomly generating score thresholds results in a much larger number of discrete scenarios presented across subjects (699, compared to the 9 scenarios presented in the primary study). The resulting rich variation in the probabilities of matriculation, when combined with orthogonal variation in correlation status, is useful for estimating our models of interest. In addition, this design allows us to conduct a new type of reduced-form test: assessing whether subjects *fail to react* to the correlation treatment when they should, in contrast to the primary experiment's test of whether subjects *react* to the correlation treatment when they should not.

We proceed with an explanation of the full contents of the experiment, summarized only briefly in the places where the set-up closely matches that in the primary experiment. Screenshots of the full experiment are available in the Materials Appendix.

After providing informed consent and completing a CAPTCHA task, subjects were presented with the set-up of the school-choice scenarios. The decision environment remained the same as in the primary study: subjects submitted an ROL listing two of three colleges, these colleges made admissions decisions by comparing randomly drawn test scores to the posted minimum test scores, and the college where the student was admitted in a randomly drawn round determined the bonus. As in the primary experiment, test scores were drawn from uniform distributions over the integers 0 to 99; bonuses for enrollment in the three colleges were \$10, \$5, or \$2.50; and subjects were randomly assigned to a single-test (correlated) treatment condition or a multiple-independent-test (independent) treatment condition. While we continued to elicit ROLs using the same interface, in this experiment we restricted choices to our focal ROLs of interest ($(A \succ B)$ or $(A \succ C)$).[27]

As in the primary experiment, after reading the initial explanation of the school-choice simulation, all subjects participated in a practice round. In this practice round, subjects faced an example set of three colleges, submitted their ROL, then saw the results. The results screen included their simulated test scores, the college where they were admitted, and the bonus they would earn if this were the incentivized scenario. After the practice round, subjects faced a brief comprehension test. They were quizzed on how the simulations affect their bonus payment and on how test scores are determined. The test score question directly assessed if subjects could correctly identify their treatment assignment. Subjects who answer these questions incorrectly were not eligible to proceed with the experiment.

After these preliminaries, subjects faced two modules of 10 school-choice scenarios. The first 10 decisions were made under their initial correlated or independent treatment assignment. After those 10 decisions were completed, we explained that we would simulate test scores in a different manner for the final 10. We then presented the explanation of the treatment condition that the subject had not yet faced. Before proceeding to the final module, subjects again faced a question testing their understanding of how test scores would be determined. Subjects who incorrectly answered that question were not eligible to continue with the experiment, and subjects who correctly answered it proceeded to the final module. Upon completing this module, the experiment concluded by simulating and presenting the results for the scenario that was randomly selected for payment.

[27]This design decisions arose from two considerations. First, because our interest is in the relative preference between these two ROLs, forcing respondents to directly express that preference is useful for statistical power. Second, even the limited rates of dominated ROLs chosen in the primary experiment suggest complications for structural estimation. Our behavioral model of interest provides an explanation for the most common "mistake" in our data: choosing the aggressive ROL when one should choose the diversified ROL. We do not believe correlation neglect is the explanation for other, rarer, mistakes. If the choice of dominated strategies comes from biases that fall outside of our model of correlation neglect, then fitting a structural model to ROL preferences that include choices of dominated strategies faces problems with model misspecification. Fitting a structural model to preferences expressed only over the diversified and aggressive strategies can be done with sounder conceptual footing, and in this binary-choice framework any residual biases can better be incorporated into the random utility term inherent in our logit framework.

The 20 scenarios presented to subjects feature admissions thresholds that were independently and randomly generated according to the following structure. First, we randomly drew the admissions threshold for College A (denoted c_A) uniformly from values 50-95. Next, we randomly drew the admissions threshold for College B (denoted c_B) uniformly between $max(2c_A - 100, 50)$ and 95. Finally, we randomly drew the admissions threshold for College C between $max(2c_B - 110, 0)$ and $min(95, 2c_B - c_A + 10)$. At all stages we rounded thresholds to the nearest 5 to avoid concerns of left digit bias. This sampling pattern guarantees that a risk-neutral expected-utility maximizer (with or without correlation neglect) would include College A in their optimal ROL (justifying our focus on $(A \succ B)$ and $(A \succ C)$). Maintaining the assumption of risk neutrality, it also ensures that every choice between $(A \succ B)$ and $(A \succ C)$ would be close to marginal in the correlated treatment arm for some value of an Enke–Zimmermann–like χ parameter drawn from the unit interval (see Section 5.2.3 for this model). Data from marginal scenarios maximally contribute to statistical power in our estimation framework, and thus by presenting scenarios that are marginal across the natural range of χ we ensure that our experiment contains a reasonable number of power-maximizing observations for the (initially unknown) value of χ that is held in the population.

5.1.2. *Preregistration.* We fully preregistered our hypotheses, analysis plan, target sample size, and sample inclusion rules prior to the beginning of data collection. We followed this preregistration plan precisely. Our preregistration is archived on aspredicted.org, and is included in the Materials Appendix for ease of reference.

5.1.3. *Deployment of Structural Experiment.* Our experiment was deployed on the Prolific experimental platform in August of 2022. On average, participation in the experiment took 11 minutes. All subjects received a $1.50 fixed payment for participation and earned an average bonus of $3.58 from the incentivized school-choice scenario. 165 subjects completed the experiment,[28] with an average age of 27 years and a 61-39 male/female gender distribution.

5.2. **Results of the Structural Experiment.**

5.2.1. *Initial Reduced-Form Examination.* Table 5 presents OLS regressions estimating cross-treatment-arm differences in the rate of choosing the aggressive ROL over the diversified ROL.

As a baseline, the dependent variable for analysis in the first column is an indicator of whether a risk-neutral expected-utility maximizer would weakly prefer the aggressive strategy. These results indicate that such a decision-maker would face large cross-treatment-arm differences: he would weakly prefer the aggressive strategy 21.9 percentage points ($s.e. = 1.7$, $p < 0.001$) less often in the correlated arm than the independent arm.

In contrast, subjects in our experiment showed very little response to treatment condition, on average. In column two of the table, the dependent variable is an indicator of whether the subject

[28]We slightly overshot our preregistered target sample size of 150 due to imperfectly forecasting the rate at which subjects would successfully pass comprehension checks. The fact that this resulted in the same sample size as the primary experiment is a coincidence.

actually chose the aggressive strategy. The results indicate that our subjects chose the aggressive strategy 3.2 percentage points ($s.e. = 1.8$, $p = 0.077$) less often in the correlated arm than the treatment arm. Similar estimates arise in columns 3-5, which include fixed effects for each discrete scenario presented, each subject, or both.

These results suggest underreaction to the differences across correlated and independent environments. Taken in isolation, this serves as a reduced-form test of a central implication of correlation neglect. While our normative benchmark in this analysis includes an assumption of risk neutrality, the more structural analyses that follows will drop this requirement and infer risk preferences from subjects' choices.

5.2.2. *Estimating Models of Sophisticated Agents.* We begin our structural analysis by examining our ability to rationalize ROL choices while assuming sophisticated assessments of probabilities. We do so by estimating McFadden-style logit random-utility models while using objective probabilities as predictors.

In this model, each individual i facing school-choice decision number j chooses which of the two possible ROLs r to submit. They submit the ROL that maximizes their utility:

$$U_{rij} = u_A p_A(r)_{ij} + u_B p_B(r)_{ij} + u_C p_C(r)_{ij} + \epsilon_{rij}. \tag{2}$$

In this equation, for each $x \in \{A, B, C\}$, u_x denotes the utility associated with matching to college x and $p_x(r)_{ij}$ denotes the probability of matching to college x conditional on submitting ROL r, taking as given the admissions thresholds presented in person i's j^{th} scenario. To conserve notation, moving forward we will suppress the i and j subscripts on these probability terms, although these terms should always be understood to take into account the specific admissions thresholds presented to a subject in a given scenario. Let $p(r)$ denote the vector $[p_A(r), p_B(r), p_C(r)]$ and let u denote the vector $[u_A, u_B, u_C]$. Let ϵ_{rij} denote a random utility perturbation drawn from the standard Gumbel distribution.

This utility framework features location and scale normalization (which are without loss). The location normalization is our assumption that the utility of remaining unmatched is zero (which explains the absence of a $p \cdot u$ term for this possibility in Equation 2). Scale normalization is achieved by the fixed variance of the random-utility error term. With this fixed variance, the magnitude of utility coefficients can be interpreted as a measure of the relative importance of the deterministic utility component ($\sum_{x \in \{A,B,C\}} p_x(r) u_x$) relative to the random utility component (ϵ_{rij})

The model expressed in Equation 2 is a direct application of the standard McFadden-style logit-utility framework. Within that framework, the $p(r)$ terms are the observed independent variables and the u terms are the coefficients on the independent variables to be estimated. Because subjects are choosing between the ROLs $(A \succ B)$ and $(A \succ C)$, and because the probability of matching to College A is the same in both of these ROLs, the term $u_A p_A(r)$ is constant across all available options in each decision and thus u_A is unidentified. However, the randomly generated admission thresholds result in rich variation in $p_B(r)$ and $p_C(r)$ that allows us to directly estimate u_B and

u_C. Because College B provides a payoff of \$5 and College C provides a payoff of \$2.50, we may directly test for risk aversion by testing whether $\frac{u_B}{u_C} < 2$.

In this framework, it will be useful to specify whether the probabilities $p(r)$ have been calculated assuming the use of a single test score (as in the correlated treatment arm) or multiple independent test scores (as in the independent treatment arm). We denote this with superscripts S (for single) and M (for multiple). This notation is useful for specifying the central predictions of both correlation neglect and sophistication. Focusing first on sophistication, the model to estimate is:

$$U_{rij} = \sum_{x \in \{A,B,C\}} \Big(\beta_x p_x^M(r) \times I(ind.\ arm) + \beta_x p_x^S(r) \times I(cor.\ arm) \Big) + \epsilon_{rij}. \tag{3}$$

If we set β_x equal to u_x for all x, then equation 3 is formally equivalent to equation 2. This presentation merely makes explicit that the true probability vector $p(r)$ is equal to $p^M(r)$ in the independent treatment arm (referenced with the indicator variable $I(ind.\ arm)$) and is equal to $p^S(r)$ in the correlated treatment arm (referenced with the indicator variable $I(cor.\ arm)$). It additionally imposes a constraint implied by sophistication: that for each $x \in \{A, B, C\}$, the coefficients on independent variables $p_x^M(r) \times I(ind.\ arm)$ and $p_x^S(r) \times I(cor.\ arm)$ are equal to each other (because both are u_x).

The first column of estimates in Table 6 presents our estimates of this model. The top panel presents the coefficients of the logit model, estimated with the sophistication constraint described above. The bottom panel summarizes the estimates of the structural parameters for ease of comparison across columns. As we see, estimates of the sophisticated model using all data generate an estimate of u_B of 10.3 ($s.e. = 0.9$) and an estimate of u_C of 5.4 ($s.e. = 0.6$). While the magnitude of these estimates viewed in isolation is complicated to interpret, the ratio of u_B to u_C (1.9, $s.e. = 0.1$) has a clear interpretation as a measure of risk tolerance. This ratio is slightly below, and not statistically different from, the value it would take if the college yielding a \$5 bonus were assigned twice the value of the college yielding a \$2.50 bonus. These estimates indicate that choices are best rationalized a sophisticated model featuring risk neutrality or slight risk aversion.

While the previous subsection's finding of a lack of responsivity to the correlation condition suggests that respondents are not reacting to at least one important feature of the decision environment, this result makes clear that respondents are reacting to at least some other important features, at least on average.

In the next two columns of Table 6, we begin to test for imperfections in the sophisticated model. If subjects correctly calculate probabilities under both treatments and use them as modeled, we should recover the same utility structure when estimating the model in either subsample. We see in the table, however, that this does not occur. The sophisticated model generates statistically distinguishable estimates across the two treatment arms, with the differences suggesting meaningfully different preferences. Using data only from the independent arm, we infer that the ratio of u_B to u_C is 1.7 ($s.e. = 0.1$). This value is significantly below 2 ($p < 0.001$), and thus indicates statistically detectable risk aversion. In contrast, using data only from the correlated arm, we infer that the ratio of u_B to u_C is 2.5 ($s.e. = 0.2$). This value is significantly above 2 ($p < 0.001$), and thus indicates statistically detectable risk seeking. This discrepancy is a demonstration of misspecification in the sophisticated model and confirms a simple prediction of correlation neglect. Because correlation-neglectful subjects overestimate the relevant probabilities in the presence of correlation, rationalizing their taste for pursuing those probabilities requires attributing greater tolerance of risk when the taste cannot be attributed to miscalculation. In addition, the smaller magnitude of coefficients in the correlated arm indicates a greater role of the random utility term in explaining choices under correlation. In short, the best-fit sophisticated model provides a better deterministic rationalization of the data in the independent arm than the correlated arm—another prediction of correlation neglect.

5.2.3. *Estimating a Model with Correlation Neglect.* Having examined models estimated under the constraints of sophistication, we now estimate models with those constraints removed to examine the ways in which they are violated. In the right segment of Table 6 we report results from estimating a logit utility model of the form:

$$U_{rij} = \sum_{x \in \{A,B,C\}} \left(\left[\beta_x^M p_x^M(r) + \beta_x^S p_x^S(r) \right] \times I(ind.\ arm) + \left[\gamma_x^M p_x^M(r) + \gamma_x^S p_x^S(r) \right] \times I(cor.\ arm) \right) + \epsilon_{rij}. \tag{4}$$

In this framework, the β terms measure the responses to both $p^M(r)$ and $p^S(r)$ in the independent treatment arm and the γ terms measure the response to both $p^M(r)$ and $p^S(r)$ in the correlated treatment arm.

In the first column of estimates presented in the "Unconstrained Model" portion of Table 6 we see an informative pattern of results.

Examining the first four rows, which summarize the coefficients on the different probabilities when estimated within the independent arm, we see large and statistically significant coefficients on the $p^M(r)$ terms—i.e. the *correct* probabilities for calculations in this treatment arm. The coefficients on the $p^S(r)$ terms—i.e., probabilities that would be correct in the correlated arm, but which are not correct here—are small in magnitude and statistically indistinguishable from zero. These results are consistent with choices being guided by the appropriate probabilities in the independent treatment arm.[29]

Examining the next four rows, which summarize the coefficients on the different probabilities when estimated within the correlated arm, we again see large and statistically significant coefficients on the $p^M(r)$ terms—i.e. the *incorrect* probabilities for calculations in this treatment arm. The coefficients on the $p^S(r)$ terms—i.e., the correct probabilities in this treatment arm—are small in magnitude and statistically indistinguishable from zero. These results are consistent with choices being guided by inappropriate probabilities in the correlated treatment arm, and specifically those that would be used if the subject neglected the presence of correlation.

[29]Notably, this finding rules out the concern that cross-arm differences are driven by "independence neglect:" incorrectly assuming correlation when assuming independence is appropriate.

A model of partial correlation neglect puts structure on the predicted values of these coefficients that may be used to estimate utility parameters. These predicted values are reported on the left side of the "Unconstrained Model" portion of Table 6 and are summarized here. A correlation neglectful agent behaves like a sophisticate in the absence of correlation, and thus the interpretation of coefficients aligns with that in the sophisticated model in the independent arm. This means that the coefficients on probabilities $p_B^M(r)$ and $p_C^M(r)$ estimate u_B and u_C, respectively, and the coefficients on probabilities $p_B^S(r)$ and $p_C^S(r)$ should be zero. Following the modeling approach of Enke and Zimmermann (2019), we assume that a partially correlation-neglectful agent applies probabilities $(1-\chi)p^S(r)+\chi p^M(r)$ in the correlated arm. If $\chi = 0$, this corresponds to correctly applying $p^S(r)$ and thus facing no neglect; if $\chi = 1$, this corresponds to incorrectly applying $p^M(r)$ and thus facing full neglect; if $\chi \in (0,1)$, this corresponds to applying an intermediate probability estimate that is biased towards the calculation that would be made assuming independence.

In the lower panel of Table 6 we provide minimum-distance estimates of χ, u_B, and u_C, all extracted from our logit regression coefficients. Our estimate of χ is 1.0 ($s.e. = 0.1$), providing a reasonably precise indication of full correlation neglect. The remaining utility parameters suggest the presence of risk aversion, with a ratio of u_B to u_C of 1.7 ($s.e. = 0.1$).

The final column of Table 6 reproduces these analyses while restricting the estimation sample to only the first module seen by each subject. While one might have worried that the lack of response to correlation could derive from failure to notice the change in treatment assignment occurring half way through the experiment, the fact that this restriction has little effect on our model estimates indicates that this cannot drive our results.[30]

5.3. **Summary and Interpretation.** In this experiment, we provide two additional tests of correlation neglect. These show that when our correlation treatment is assigned statistically independently from admissions thresholds, subjects underreact to treatment assignment, and that ROL choices are well predicted by the probabilities one would calculate assuming independence (whether in the independent or correlated decision environment). In addition, when estimating the fully preregistered structural model that our experiment was designed to identify, we find that the data are best fit by a model with precisely estimated full correlation neglect. These findings strongly support our proposal that correlation neglect functions as an important driver of mistakes in this domain.

6. Discussion

In this paper, we have noted that correlation neglect offers a natural explanation for some of the difficulties observed in a common class of school-choice problems: those in which rankings are submitted prior to essential test scores being determined. We have followed in the tradition of work such as Chen and Sönmez (2006) in using controlled lab experiments to directly assess students'

[30]This worry is further assuaged by our choice to include subjects who correctly identified their treatment condition in the comprehension checks in our data, and our finding that the results of Table 5 are also closely reproduced when estimated from only module 1 data (see Appendix Table A5).

response to school-assignment mechanisms. We note, however, that these issues are not unique to centralized markets—similar considerations are relevant in decentralized school admissions problems as well.

Difficulties in comprehension induced by correlation are broadly relevant for market designers. The systems we highlighted in Section 1 offer somewhat extreme examples of the forecasting challenges that we have described, but some degree of these challenges are ubiquitously present. Whether due to institutional constraints or scarce consideration time, students often have to choose a comparatively small set of schools to ultimately apply to or rank. Even in cases where all inputs to evaluation are known, students often harbor some uncertainty about how their profile will be evaluated. The fact that programs often have some agreement on the evaluation process results in correlation. In such environments, students' approach to the matching process might be meaningfully suboptimal not due to a failure to optimally rank-order the schools to which they apply (the typical mistake of interest in behavioral matching papers, see Hakimov and Kübler (2020) for a recent review), but rather due to applying to the wrong set of schools in the first place. And indeed, forming an unwise application list can be a substantially more costly mistake, since the primary risk induced is not matching to the wrong school (as arises from misordering preferences submitted to the deferred acceptance algorithm), but rather failing to match at all.

A common reaction to the results of this paper is to note that correlation neglect only becomes relevant due to the imposition of constraints on the length of application lists. As is well known within the market design literature, imposing these constraints comes with costs even among sophisticated agents (see, e.g., Haeringer and Klijn 2009, Calsamiglia et al. 2010). Perhaps most notably, constrained application lists eliminate the strategy-proofness of the deferred acceptance algorithm, leading most of the theoretical literature to model application lists as unconstrained. If these theoretical insights led to constraints only rarely being imposed, our study would simply provide another reason to avoid an already-avoided ingredient for market design. However, we note that in practice, constraints on the number of applications are nearly ubiquitous in school-choice applications. To illustrate, Pathak and Sönmez (2013) summarize 70 school admissions reforms occurring between 1999 and 2012 and find that all but 3 resulted in the implementation of a constrained system.[31] In light of this practical reality, we believe that behavioral factors interacting with constraints must be understood and accommodated.

Beyond its relevance for assessing constraints, the presence of correlation neglect directly influences several other practical considerations of market design. One immediate consequence is a connection between the literature on priority-tie-breaking in school choice to the discussion on protecting unsophisticated agents (e.g., Pathak and Sönmez 2008, Hassidim et al. 2017a, Rees-Jones 2017). Tie-breaking methods have been studied extensively, especially comparing the use of a single common lottery with multiple independent lotteries (e.g., Abdulkadiroğlu et al. 2009,

[31]The most common system adopted was deferred-acceptance constrained to lists of 3 schools—i.e., a system nearly as constrained as those in our experiment.

Ashlagi et al. 2019).[32] Our results imply that, when applications are restricted or costly, there may exist a tension between efficiency (which may be improved by the use of a single lottery) and the desire to protect unsophisticated agents (which calls for using multiple independent lotteries). When concerns about protecting the unsophisticated become central, the use of multiple tie-breaking rules is recommended.

In summary, and to conclude, the specific failures in reasoning induced by correlation neglect directly interplay with crucial technical aspects of market design. Based on the magnitude of effects observed in our study, combined with the substantial prevalence of suboptimal choices seen in the matching systems reviewed in Section 1, we believe that integration and accommodation of correlation neglect into our frameworks for market design can be of substantial benefit to the millions of students who interact with these systems.

References

ABDULKADIROĞLU, A., P. A. PATHAK, AND A. E. ROTH (2009): "Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match," American Economic Review, 99, 1954–78.

AGRANOV, M. AND P. ORTOLEVA (2017): "Stochastic Choice and Preferences for Randomization," Journal of Political Economy, 125, 40–68.

AJAYI, K. F. (2013): "School Choice and Educational Mobility: Lessons from Secondary School Applications in Ghana," Mimeo.

AJAYI, K. F., W. H. FRIEDMAN, AND A. M. LUCAS (2020): "When Information is Not Enough: Evidence from a Centralized School Choice System," Working Paper 27887, National Bureau of Economic Research.

ASHLAGI, I., A. NIKZAD, AND A. ROMM (2019): "Assigning More Students to their Top Choices: A Comparison of Tie-Breaking Rules," Games and Economic Behavior, 115, 167 – 187.

BASTECK, C. AND M. MANTOVANI (2018): "Cognitive Ability and Games of School Choice," Games and Economic Behavior, 109, 156 – 183.

BORDALO, P., N. GENNAIOLI, AND A. SHLEIFER (2012): "Salience Theory of Choice Under Risk," Quarterly Journal of Economics, 127, 1243–1285.

BROECKE, S. (2012): "University Selectivity and Earnings: Evidence from UK Data on Applications and Admissions to University," Economics of Education Review, 31, 96–107.

BUSHONG, B., M. RABIN, AND J. SCHWARTZSTEIN (2020): "A Model of Relative Thinking," The Review of Economic Studies, 88, 162–191.

CALSAMIGLIA, C., G. HAERINGER, AND F. KLIJN (2010): "Constrained School Choice: An Experimental Study," American Economic Review, 100, 1860–74.

CASON, T. AND C. PLOTT (2014): "Misconceptions and game form recognition: Challenges to theories of revealed preference and framing," Journal of Political Economy, 122, 1235–1270.

CERREIA-VIOGLIO, S., D. DILLENBERGER, P. ORTOLEVA, AND G. RIELLA (2019): "Deliberately Stochastic," American Economic Review, 109, 2425–45.

CHADE, H. AND L. SMITH (2006): "Simultaneous Search," Econometrica, 74, 1293–1307.

[32] While this literature focused on centralized mechanisms, admission decisions are also independent in decentralized markets where oversubscribed schools use independent (school-specific) lotteries to break priority ties between applicants (e.g., Dobbie and Fryer 2011). Similarly, if all schools use a single lottery or exam to break priority ties, admission decisions are correlated.

CHEN, Y., P. CRAMTON, J. A. LIST, AND A. OCKENFELS (2020): "Market Design, Human Behavior, and Management," Management Science, 67, 5317–5348.

CHEN, Y. AND T. SÖNMEZ (2006): "School Choice: An Experimental Study," Journal of Economic Theory, 127, 202–231.

DARGNIES, M.-P., R. HAKIMOV, AND D. KÜBLER (2019): "Self-Confidence and Unraveling in Matching Markets," Management Science, 65, 5603–5618.

DING, T. AND A. SCHOTTER (2017): "Matching and Chatting: An Experimental Study of the Impact of Network Communication on School-Matching Mechanisms," Games and Economic Behavior, 103, 94–115.

DOBBIE, W. AND J. FRYER, ROLAND G. (2011): "Are High-Quality Schools Enough to Increase Achievement among the Poor? Evidence from the Harlem Children's Zone," American Economic Journal: Applied Economics, 3, 158–87.

DREYFUSS, B., O. HEFFETZ, AND M. RABIN (2021): "Expectations-Based Loss Aversion May Help Explain Seemingly Dominated Choices in Strategy-Proof Mechanisms," American Economic Journal: Microeconomics, forthcoming.

DUBINS, L. E. AND D. A. FREEDMAN (1981): "Machiavelli and the Gale-Shapley Algorithm," The American Mathematical Monthly, 88, 485–494.

DWENGER, N., D. KÜBLER, AND G. WEIZSÄCKER (2018): "Flipping a Coin: Evidence from University Applications," Journal of Public Economics, 167, 240–250.

ENKE, B. AND F. ZIMMERMANN (2019): "Correlation Neglect in Belief Formation," The Review of Economic Studies, 86, 313–332.

ESPONDA, I. AND E. VESPA (2014): "Hypothetical Thinking and Information Extraction in the Laboratory," American Economic Journal: Microeconomics, 6, 180–202.

——— (2019): "Contingent Thinking and the Sure-Thing Principle: Revisiting Classic Anomalies in the Laboratory," Mimeo.

EYSTER, E. AND G. WEIZSÄCKER (2016): "Correlation Neglect in Portfolio Choice: Lab Evidence," SSRN WP 2914526.

FEATHERSTONE, C. R. AND M. NIEDERLE (2016): "Boston versus Deferred Acceptance in an Interim Setting: An Experimental Investigation," Games and Economic Behavior, 100, 353 – 375.

FIEDLER, K. AND P. JUSLIN (2006): "Taking the Interface Between Mind and Environment Seriously," Information Sampling and Adaptive Cognition, 3–29.

GALE, D. AND L. SHAPLEY (1962): "College Admissions and Stability of Marriage," American Mathematical Monthly, 69, 9–15.

GUILLEN, P. AND R. HAKIMOV (2017): "Not Quite the Best Response: Truth-telling, Strategy-proof Matching, and the Manipulation of Others," Experimental Economics, 20, 670–686.

——— (2018): "The Effectiveness of Top-down Advice in Strategy-proof Mechanisms: A Field Experiment," European Economic Review, 101, 505–511.

HAERINGER, G. AND F. KLIJN (2009): "Constrained School Choice," Journal of Economic Theory, 144, 1921 – 1947.

HAKIMOV, R. AND D. KÜBLER (2020): "Experiments on Centralized School Choice and College Admissions: A Survey," Experimental Economics.

HASSIDIM, A., D. MARCIANO, A. ROMM, AND R. I. SHORRER (2017a): "The Mechanism Is Truthful, Why Aren't You?" American Economic Review, 107, 220–24.

HASSIDIM, A., A. ROMM, AND R. I. SHORRER (2017b): "Redesigning the Israeli Psychology Master's Match," American Economic Review, 107, 205–09.

——— (2021): "The Limits of Incentives in Economic Matching Procedures," Management Science, 67, 951–963.

KÖSZEGI, B. AND A. SZEIDL (2012): "A Model of Focusing in Economic Choice," The Quarterly Journal of Economics, 128, 53–104.

KOUTOUT, K., A. DUSTAN, M. VAN DER LINDEN, AND M. WOODERS (2021): "Mechanism Performance Under Strategy Advice and Sub-Optimal Play: A School Choice Experiment," Journal of Behavioral and Experimental Economics, 94, 101755.

LEVIN, D., J. PECK, AND A. IVANOV (2016): "Separating Bayesian Updating from Non-Probabilistic Reasoning: An Experimental Investigation," American Economic Journal: Microeconomics, 8, 39–60.

LEVY, G. AND R. RAZIN (2015): "Correlation Neglect, Voting Behavior, and Information Aggregation," American Economic Review, 105, 1634–45.

LI, S. (2017): "Obviously Strategy-Proof Mechanisms," American Economic Review, 107, 3257–87.

LUCAS, A. M. AND I. M. MBITI (2012): "The Determinants and Consequences of School Choice Errors in Kenya," American Economic Review, 102, 283–88.

MARTÍNEZ-MARQUINA, A., M. NIEDERLE, AND E. VESPA (2019): "Failures in Contingent Reasoning: The Role of Uncertainty," American Economic Review, 109, 3437–3474.

PAN, S. (2019): "The Instability of Matching with Overconfident Agents," Games and Economic Behavior, 113, 396 – 415.

PATHAK, P. A. AND T. SÖNMEZ (2008): "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism," American Economic Review, 98, 1636–52.

PATHAK, P. A. AND T. SÖNMEZ (2013): "School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation," American Economic Review, 103, 80–106.

PERANSON, J. (2019): "Design and Implementation of the Genetic Counseling Admissions Match," in Proceedings of MATCH-UP 2019, 5th International Workshop on Matching Under Preferences, Match-UP, 21.

RAVEN, J. AND J. RAVEN (2003): Raven Progressive Matrices, Boston, MA: Springer US, 223–237.

REES-JONES, A. (2017): "Mistaken Play in the Deferred Acceptance Algorithm: Implications for Positive Assortative Matching," American Economic Review, 107, 225–29.

——— (2018): "Suboptimal Behavior in Strategy-Proof Mechanisms: Evidence from the Residency Match," Games and Economic Behavior, 108, 317 – 330.

REES-JONES, A. AND S. SKOWRONEK (2018): "An Experimental Investigation of Preference Misrepresentation in the Residency Match," Proceedings of the National Academy of Sciences, 115, 11471–11476.

ROTH, A. E. (1982): "The Economics of Matching: Stability and Incentives," Mathematics of Operations Research, 7, 617–628.

SHORRER, R. I. (2019): "Simultaneous Search: Beyond Independent Successes," Mimeo.

SHORRER, R. I. AND S. SÓVÁGÓ (2018): "Obvious Mistakes in a Strategically Simple College Admissions Environment: Causes and Consequences," SSRN WP 2993538.

SOMERVILLE, J. (2019): "Choice-Set-Dependent Preferences in Consumer Choice: An Experimental Test," Mimeo.

UCAS (2011): "Admissions Process Review Consultation," Available at https://www.ucas.com/file/956/download?token=y8EovXLo.

——— (2012): "Admissions Process Review Findings and Recommendations," Available at https://www.ucas.com/file/776/download?token=6U$_C IbPI$.

WYNESS, G. (2016): "Predicted Grades: Accuracy and Impact," University and College Union Report, 1–22.

FIGURE 1. Application Strategies Across All Scenarios.

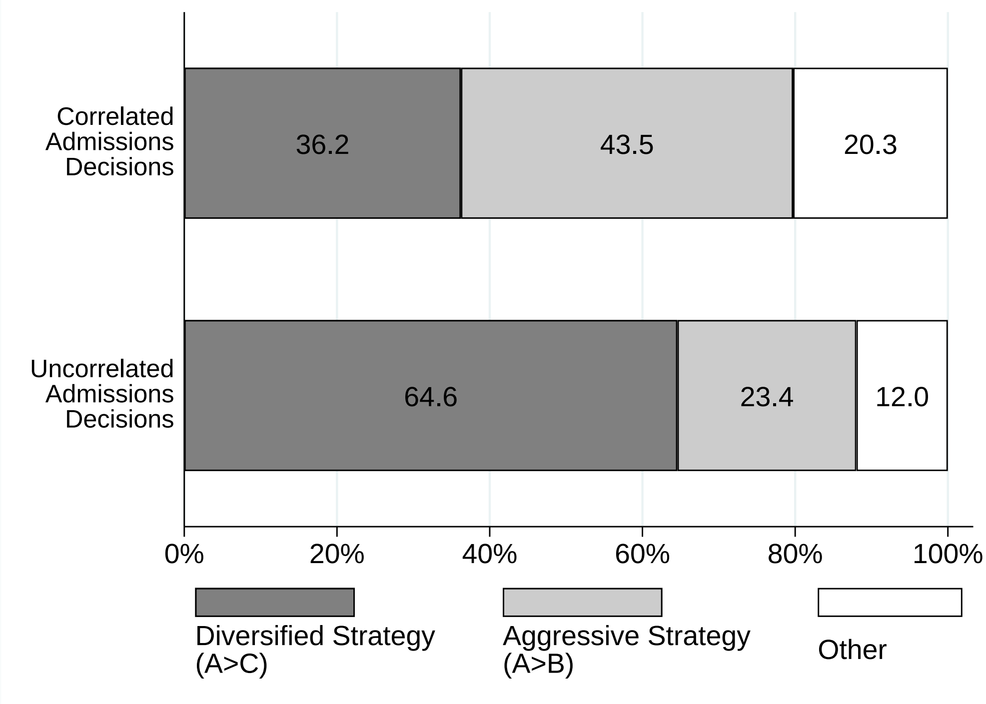


FIGURE 2. Consistency Between Application Strategies and Lottery Choices.

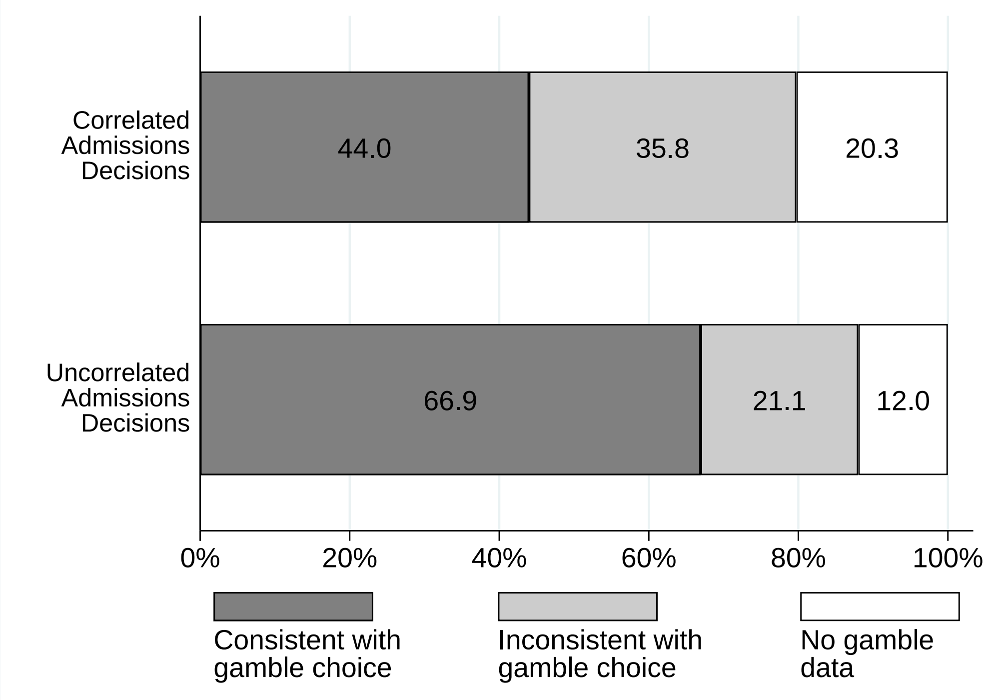
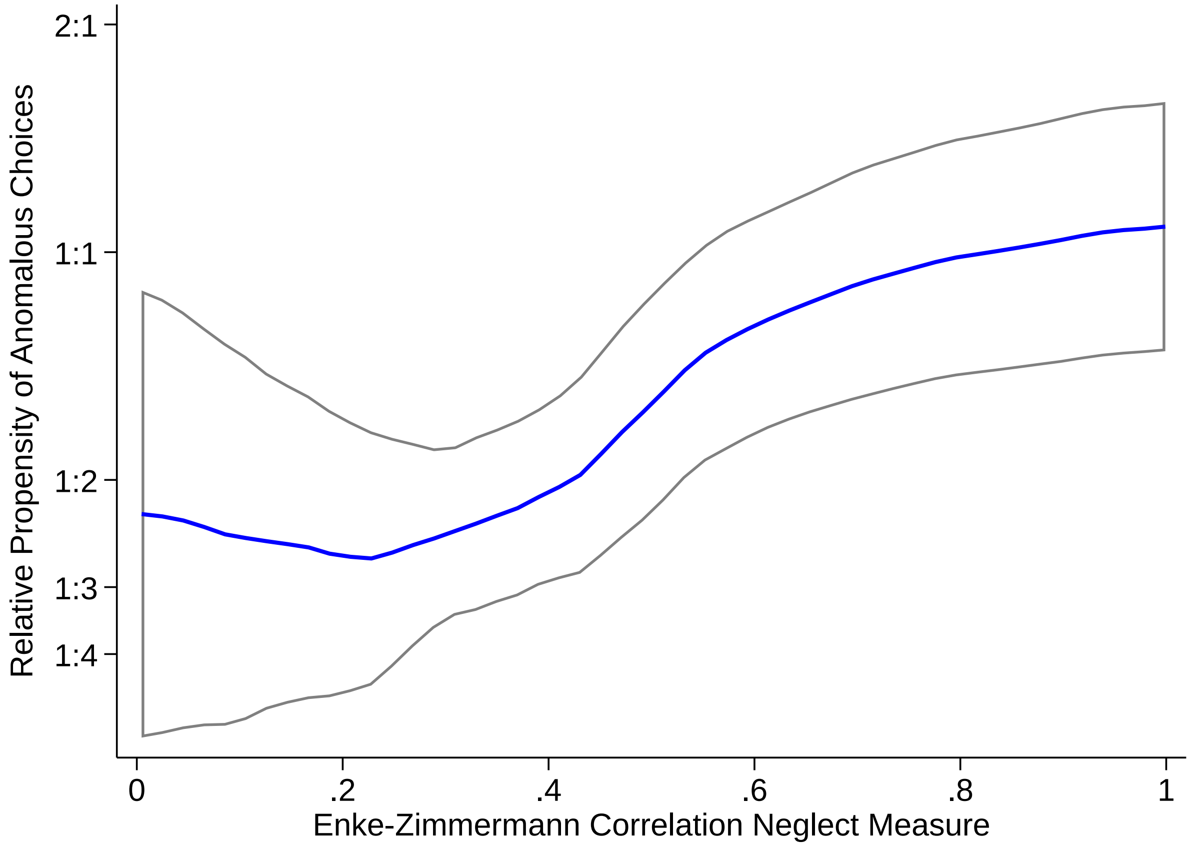


FIGURE 3. Predicting Application Mistakes with the Enke-Zimmermann Measure.



Notes: This figure presents a local-polynomial regression of the relative propensity of our target preference reversal on the Enke-Zimmermann measure of correlation neglect (restricted to the unit inverval). To illustrate the interpretation of the y-axis, note that "2:1" indicates that among the nine scenarios, the subject made a correlation-neglectful preference reversal ($(A \succ B)$ under correlation and $(A \succ C)$ under independence) twice per every optimal response ($(A \succ C)$ in both framings). Bandwidth: 0.2. Kernel: Epanechnikov. Degree: 0. Confidence level: 95%. Number of observations: 94.

TABLE 1. Parameters of All Scenarios.

Scenario		*Required Test Score* A	B	C	*Consequence of ROLs* (A≻B)	(A≻C)
1.	C	50	45	0	$\underbrace{(\$10, 0.5; \$5, 0.05; \$0, 0.45)}_{\text{EV=\$5.25}}$	$\underbrace{(\$10, 0.5; \$2.5, 0.5)}_{\text{EV=\$6.25}}$
	U	50	90	0		
2.	C	50	45	10	$\underbrace{(\$10, 0.5; \$5, 0.05; \$0, 0.45)}_{\text{EV=\$5.25}}$	$\underbrace{(\$10, 0.5; \$2.5, 0.4; \$0, 0.1)}_{\text{EV=\$6.00}}$
	U	50	90	20		
3.	C	50	20	0	$\underbrace{(\$10, 0.5; \$5, 0.3; \$0, 0.2)}_{\text{EV=\$6.5}}$	$\underbrace{(\$10, 0.5; \$2.5, 0.5)}_{\text{EV=\$6.25}}$
	U	50	40	0		
4.	C	50	20	10	$\underbrace{(\$10, 0.5; \$5, 0.3; \$0, 0.2)}_{\text{EV=\$6.5}}$	$\underbrace{(\$10, 0.5; \$2.5, 0.4; \$0, 0.1)}_{\text{EV=\$6.00}}$
	U	50	40	20		
5.	C	50	55	0	$\underbrace{(\$10, 0.5; \$0, 0.5)}_{\text{EV=\$5.00}}$	$\underbrace{(\$10, 0.5; \$2.5, 0.5)}_{\text{EV=\$6.25}}$
	U	50	100	0		
6.	C	75	60	0	$\underbrace{(\$10, 0.25; \$5, 0.15; \$0, 0.6)}_{\text{EV=\$3.25}}$	$\underbrace{(\$10, 0.25; \$2.5, 0.75)}_{\text{EV=\$4.375}}$
	U	75	80	0		
7.	C	75	60	30	$\underbrace{(\$10, 0.25; \$5, 0.15; \$0, 0.6)}_{\text{EV=\$3.25}}$	$\underbrace{(\$10, 0.25; \$2.5, 0.45; \$0, 0.3)}_{\text{EV=\$3.625}}$
	U	75	80	40		
8.	C	80	60	0	$\underbrace{(\$10, 0.2; \$5, 0.2; \$0, 0.6)}_{\text{EV=\$3.00}}$	$\underbrace{(\$10, 0.2; \$2.5, 0.8)}_{\text{EV=\$4.00}}$
	U	80	75	0		
9.	C	80	60	40	$\underbrace{(\$10, 0.2; \$5, 0.2; \$0, 0.6)}_{\text{EV=\$3.00}}$	$\underbrace{(\$10, 0.2; \$2.5, 0.4, \$0, 0.4)}_{\text{EV=\$3.00}}$
	U	80	75	50		

Notes: This table summarizes the nine "matched pairs" of scenarios presented in the primary experiment. Each numbered pair of rows indicates a given scenario pair. Row C presents the test-score thresholds presented in the correlated-admissions module, whereas row U presents the test-score thresholds presented in the uncorrelated-admissions module. The last two columns of the table present the lotteries induced by the two focal admissions strategies. Within the parenthesis, we present the monetary outcomes and their probabilities. Below each induced lottery, we present the expected value.

TABLE 2. Choices in the Correlated and Uncorrelated Settings.

Scenario		Rank-Order List $(A \succ B)$	$(A \succ C)$	$(B \succ C)$	Other	Test of Equality (p-values) Full Dist.	$(A \succ B)$	$(A \succ C)$
1.	C: (50, 45, 0)	48.5	44.9	4.2	2.4	0.01**	<0.01***	<0.01***
	U: (50, 90, 0)	10.9	84.2	0.0	4.9			
2.	C: (50, 45, 10)	50.3	44.2	3.0	2.4	<0.01***	<0.01***	<0.01***
	U: (50, 90, 20)	10.3	87.9	0.0	1.8			
3.	C: (50, 20, 0)	74.6	18.8	6.1	0.6	<0.01***	<0.01***	<0.01***
	U: (50, 40, 0)	49.7	40.6	6.1	3.6			
4.	C: (50, 20, 10)	81.8	12.7	4.9	0.6	<0.01***	<0.01***	<0.01***
	U: (50, 40, 20)	67.9	24.9	3.6	3.6			
5.	C: (50, 55, 0)	26.7	69.1	1.2	3.0	<0.01***	<0.01***	<0.01***
	U: (50, 100, 0)	7.9	87.9	0.6	3.6			
6.	C: (75, 60, 0)	24.9	45.4	23.0	6.7	<0.01***	<0.01***	<0.01***
	U: (75, 80, 0)	12.7	76.4	3.6	7.3			
7.	C: (75, 60, 30)	30.3	38.2	25.5	6.1	<0.01***	<0.01***	<0.01***
	U: (75, 80, 40)	14.6	76.4	0.6	8.5			
8.	C: (80, 60, 0)	24.2	29.1	40.0	6.7	<0.01***	0.03**	<0.01***
	U: (80, 75, 0)	14.6	57.6	18.8	9.1			
9.	C: (80, 60, 40)	30.3	23.6	39.4	6.7	<0.01***	0.10	<0.01***
	U: (80, 75, 50)	22.4	45.5	21.2	10.9			

Notes: This table summarizes the ROLs chosen in each matched pair of scenarios. All numbers presented (with the exception of the final three columns) are percentages of responses seen within a module. Columns $(A \succ B)$, $(A \succ C)$, and $(B \succ C)$ present the fractions of subjects reporting each of those ROLs, and column "other" reports the fraction of subjects reporting one of the (clearly dominated) strategies $(B \succ A)$, $(C \succ A)$, or $(C \succ B)$. The final 3 columns present p-values associated with tests for differences across the correlated and uncorrelated presentations. The column marked "Full Dist." presents the results of Fisher's exact tests of differences in the distribution of the six possible ROLs by correlation condition. The following two columns present two-sample difference-of-proportions tests, comparing the proportion picking each of the focal strategies across correlation conditions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 3. Within-Subject Application Behavior in Matched Pairs.

Scenario	Not Influenced by Correlation		Influenced by Correlation				
	C: $(A \succ C)$ U: $(A \succ C)$	C: $(A \succ B)$ U: $(A \succ B)$	C: $(A \succ B)$ U: $(A \succ C)$	C: $(A \succ C)$ U: $(A \succ B)$	Included $(B \succ C)$	Inc. Reverse Alphabetical	Frac choosing U: $(A \succ C)$ given C: $(A \succ B)$.
1. C: (50, 45, 0) U: (50, 90, 0)	41.2	8.5	37.6	2.4	4.2	6.1	77.5
2. C: (50, 45, 10) U: (50, 90, 20)	43.6	7.9	41.2	0.6	3.0	3.6	81.9
3. C: (50, 20, 0) U: (50, 40, 0)	12.7	42.4	25.5	6.1	9.7	3.6	34.1
4. C: (50, 20, 10) U: (50, 40, 20)	7.9	60.6	14.6	4.9	7.9	4.2	17.8
5. C: (50, 55, 0) U: (50, 100, 0)	64.9	4.9	21.2	3.0	1.2	4.9	79.5
6. C: (75, 60, 0) U: (75, 80, 0)	41.8	9.7	13.3	1.8	21.8	11.5	53.7
7. C: (75, 60, 30) U: (75, 80, 40)	33.9	10.9	17.6	3.0	22.4	12.1	58.0
8. C: (80, 60, 0) U: (80, 75, 0)	22.4	10.3	8.5	1.8	45.5	11.5	35.0
9. C: (80, 60, 40) U: (80, 75, 50)	17.6	13.9	10.9	2.4	40.6	14.6	36.0
Average	31.8	18.8	21.1	2.9	17.4	8.0	52.6

Notes: This table summarizes the within-subject patterns of ROLs submitted for each matched pair of scenarios. We focus on characterizing the use of the two focal application strategies, the aggressive application strategy ($(A \succ B)$) and the diversified application strategy ($(A \succ C)$). The first two columns present the fraction of subjects pursuing the diversified or the aggressive strategies, respectively, under both correlation conditions. These are the subjects who pursue a focal strategy and are not influenced by correlation. The next two columns present the fraction of subjects who switch from one focal strategy to the other based on the correlation condition—i.e., the subjects who are influenced by correlation. The next two columns characterize the fraction of subjects for whom at last one submitted ROL was a non-focal strategy: either $(B \succ C)$ or one in which programs are submitted in reverse-alphabetical order. The final column presents the fraction of subjects who chose the diversified strategy in the uncorrelated-admissions module conditional on choosing the aggressive strategy in the correlated-admissions module.

TABLE 4. Choices in Lotteries.

Lottery question	% chose $(A \succ C)$ in lottery	% chose $(A \succ B)$ in lottery cond. on ROL responding to correlation	% chose $(A \succ C)$ in lottery cond. on $(A \succ C)$ in both ROLs
1. $(A \succ B)$: \$10 w/50%, \$5 w/5% $(A \succ C)$: \$10 w/50%, \$2.5 w/50%	92.7	10.6	92.6
2. $(A \succ B)$: \$10 w/50%, \$5 w/5% $(A \succ C)$: \$10 w/50%, \$2.5 w/40%	97.6	1.4	98.6
3. $(A \succ B)$: \$10 w/50%, \$5 w/30% $(A \succ C)$: \$10 w/50%, \$2.5 w/50%	68.5	25.0	76.2
4. $(A \succ B)$: \$10 w/50%, \$5 w/30% $(A \succ C)$: \$10 w/50%, \$2.5 w/40%	47.3	59.4	69.2
5. $(A \succ B)$: \$10 w/50% $(A \succ C)$: \$10 w/50%, \$2.5 w/50%	97.0	5.7	97.2
6. $(A \succ B)$: \$10 w/25%, \$5 w/15% $(A \succ C)$: \$10 w/25%, \$2.5 w/75%	98.2	0.0	98.6
7. $(A \succ B)$: \$10 w/25%, \$5 w/15% $(A \succ C)$: \$10 w/25%, \$2.5 w/45%	98.2	5.8	100.0
8. $(A \succ B)$: \$10 w/20%, \$5 w/20% $(A \succ C)$: \$10 w/20%, \$2.5 w/80%	97.0	11.7	97.3
9. $(A \succ B)$: \$10 w/20%, \$5 w/20% $(A \succ C)$: \$10 w/20%, \$2.5 w/40%	88.5	31.8	100.0
Average	87.2	15.2	95.0

Notes: This table summarizes the choices made over pairs of lotteries constructed to offer the same payouts as arise from each scenario's focal strategies. The first column reports the fraction of subjects choosing the lottery that arises from the diversified application strategy, illustrating that this option is overwhelmingly preferred when the consequences are made transparent. The second column shows the fraction of subjects choosing the lottery that arises from the aggressive application strategy contingent on being coded as responding to correlation in the analysis of Table 3. The third column shows the fraction of subjects choosing the lottery associated with the diversified application strategy conditional on pursuing that strategy in both correlation conditions.

TABLE 5. Assessing the Rate of Aggressive Strategies Across Treatment Arms

	Should Be Aggressive	Actually Aggressive			
Correlated Arm	-0.219***	-0.032*	-0.030	-0.032*	-0.033*
	(0.017)	(0.018)	(0.020)	(0.019)	(0.020)
Constant	0.736***	0.573***			
	(0.012)	(0.019)			
Scenario FE	No	No	Yes	No	Yes
Subject FE	No	No	No	Yes	Yes
Observations	3300	3300	3300	3300	3300

Notes: This table presents OLS regressions assessing differences in the rate of pursuing the aggressive strategy across treatment arms. As a benchmark, the first column uses an indicator of whether a risk-neutral expected utility maximizer would weakly prefer the aggressive strategy as the dependent variable. The remaining columns use an indicator of whether the subject choose the aggressive strategy. After presenting the baseline OLS regression, across columns we vary whether we include fixed effects for each discrete scenario (i.e., each potential combination of admissions thresholds), for each subject, or both. Standard errors, clustered by subject, are presented in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 6. Estimating Logit Models

			Sophisticated Model				Unconstrained Model		
			Prediction	**All Data**	**Ind. Arm**	**Cor. Arm**	**Prediction**	**All Data**	**Mod. 1**
Arm-Specific Logit-Model Estimates	**Multiple Scores (Ind. Arm)**	$p^M_B(r)$	u_B	10.3***	18.5***		u_B	19.6***	17.8***
				(0.9)	(1.5)			(2.0)	(3.1)
		$p^M_C(r)$	u_C	5.4***	10.7***		u_C	11.5***	11.7***
				(0.6)	(0.9)			(1.7)	(2.8)
		$p^S_B(r)$					0	-2.8	-3.2
								(2.0)	(2.8)
		$p^S_C(r)$					0	-0.9	-2.1
								(1.2)	(1.9)
	Single Score (Cor. Arm)	$p^M_B(r)$					χu_B	18.0***	18.3***
								(1.7)	(2.3)
		$p^M_C(r)$					χu_C	11.1***	11.5***
								(1.5)	(2.0)
		$p^S_B(r)$	u_B	10.3***		9.8***	$(1-\chi)u_B$	-1.6	-0.8
				(0.9)		(1.6)		(1.9)	(2.5)
		$p^S_C(r)$	u_C	5.4***		3.9***	$(1-\chi)u_C$	-0.6	-0.8
				(0.6)		(0.6)		(1.1)	(1.6)
Implied Parameters		u_B/u_C		1.9***	1.7***	2.5***		1.7***	1.6***
				(0.1)	(0.1)	(0.2)		(0.1)	(0.1)
		u_B		10.3***	18.5***	9.8***		18.5***	17.7***
				(0.9)	(1.5)	(1.6)		(1.4)	(2.1)
		u_C		5.4***	10.7***	3.9***		11.1***	11.4***
				(0.6)	(0.9)	(0.6)		(1.2)	(1.7)
		χ						1.0***	1.0***
								(0.1)	(0.1)
(r,i,j) Obs.				6,600	3,300	3,300		6,600	3,300
Subjects				165	165	165		165	165

Notes: This table presents estimates of McFadden logit utility models. The columns grouped under "Unconstrained Model" report estimates of the model presented in equation 4. The columns grouped under "Sophisticated Model" report estimates of the nested model presented in equation 3 that imposes constraints implied by sophistication. The first column of each group reports the theoretically predicted value of the estimated coefficients. The remaining columns in each group estimate these coefficients in different samples: using all data, using only data from the independent arm (“Ind. Arm”) or the correlated arm (“Cor. Arm”), or using data only from the first module seen (“Mod. 1”). Standard errors, clustered by subject, are reported in parentheses. The lower panel reports the structural parameter values implied by the estimates. For the “Sophisticated Model” parameters, estimates for u_B and u_C are simply the top-panel model estimates. The ratio $\frac{u_B}{u_C}$ and its standard errors are estimated with the delta method. For the “Unconstrained Model" parameters, all terms are estimated by minimizing the sum of squared distances between the predicted coefficient values and estimated valeus reported in the top panel. Standard errors are block-bootstrapped (blocked at the subject level) with 1,000 iterations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendices

A. Proofs

Proposition 1. *For any school choice environment and for any undominated ROL r, $V_s(r) \leq V_n(r)$.*

Proof. The proof proceeds by induction on the size of the ROL. The case $k = 1$ is obvious as correlation in admission decisions is irrelevant for students' subjective expected utility. For the case $k > 1$, let $r^{2:k}$ denote the continuation ROL from the second to the k-th ranked schools. Then, we have that for all $x \in \{s, n\}$,

$$V_x(r) = (1 - F(c_{r^1}))\, u\left(r^1\right) + F(c_{r^1})\, V_x(r^{2:k} \mid \text{rejected by } r^1).$$

For the neglectful type, $V_n(r^{2:k} \mid \text{rejected by } r^1) = V_n(r^{2:k})$. For the sophisticated type, $V_s(r^{2:k} \mid \text{rejected by } r^1) \leq V_s(r^{2:k})$, as the absence of information results in a first order stochastically higher distribution of outcomes (mass is reduced proportionally from all options and added to r^2).[33]

By induction, $V_n(r^{2:k}) \geq V_s(r^{2:k})$. Altogether we have that

$$V_n(r^{2:k} \mid \text{rejected by } r^1) = V_n(r^{2:k}) \geq V_s(r^{2:k}) \geq V_s(r^{2:k} \mid \text{rejected by } r^1),$$

and hence

$$V_n(r) \geq V_s(r).$$

□

Proposition 2. *For any integer k, and any decision environment where the agent is constrained to (costlessly) apply to up-to-k schools, the price of neglect for the neglectful type is bounded above*

[33] Here we use the fact that undominated ROLs are ordered according to true preferences.

by $1 - \frac{1}{k}$. Furthermore, this bound is tight—for any k, there exist school choice environments where the price of neglect is arbitrarily close to $1 - \frac{1}{k}$.

Proof. To begin, note that the optimal size-1 ROL is identical for all types, as correlation only matters when applying to multiple schools. Next, observe that the neglectful type believes that admissions decisions across schools are independent. Thus, by Theorem 1 of Chade and Smith (2006), any subjective-optimal ROL of size-k of the neglectful type includes a subjective-optimal singleton ROL, which is also an objective optimal singleton ROL by the first observation. Thus, subjective-optimal ROLs of size-k achieve at least as much experienced utility as the optimal size-1 ROL (the fact that the ROL of the neglectful type includes more schools can only improve the utility he will experience, as he will attend the best school that accepts him). Finally, by Theorem 2 of Shorrer (2019), the expected utility of a sophisticated agent from an optimal size-1 ROL is greater than or equal to $\frac{1}{k}$ of the expected utility from an optimal size-k ROL.

To see that the lower bound is tight, consider an arbitrarily small $\epsilon > 0$. For $m \in \{1, 2, ..., k - 1\}$, let $u_m := \epsilon^{-m}$ and let $c_m := 1 - \epsilon^m$, and let $u_k := \epsilon^{-k}(1 + \delta)$ and $c_k := 1 - \epsilon^k$. Let X consists of k copies of each type of school, (u_i, c_i). Then the full correlation neglectful type will choose the k copies of the most desired school, u_k, and get utility of $1 + \delta$ (see, e.g. Chade and Smith 2006). But by choosing one school of each type the expected utility is approximately k for sufficiently small ϵ and δ. □

Proposition 3. *For any constraint on the size of the ROL k, the neglectful type is at least as likely to be unassigned as the sophisticated type.*

Our leading example shows that this comparison may be strict.

Proof. The case of $k = 1$ is obvious since correlation plays no role when students can only apply to one school. Next, recall from Shorrer (2019) that options that are more selective and less desirable than other options are dominated and do not appear on an optimal ROL of a sophisticated agent.

A consequence of this statement is that when considering the sophisticated type's ROL, there is no loss in focusing on a subset of undominated alternatives $X' \subset X$ such that for any $x, y \in X'$, $u(x) > u(y) \iff c_x > c_y$.

Consider the subjective-optimal size-k ROL of the neglectful type, $r_n(k)$. Since options in $X \backslash X'$ are dominated, they can only appear on ROLs that include choices that dominate them. Thus, the least selective school on $r_n(k)$ belongs to X'. Hence,$|r_n(k) \cap (X \backslash X')| = m < k$.

Consider the choice problem where an agent needs to choose optimal ROLs of size $k - m$ from X' with the stochastic outside option of m independent lotteries, one for each $i \in r_n(k) \cap (X \backslash X')$, where the probability of realization of lottery i is $1 - c_i$ and its utility from attending is $u(i)$ (but the student still can only attend one school). The outside option is how the neglectful type subjectively perceives $r_n(k) \cap (X \backslash X')$. Note that since optimal ROLs rank schools according to desirability, in this problem, the lowest-ranked school is the least selective option in the ROL (ignoring the outside options).

We now claim that the last (i.e., $k - m$-th ranked) school on the neglectful type's ROL is associated with a weakly more selective threshold than that of the sophisticated type. Towards contradiction, assume the opposite. Then the last choice on the neglectful type's ROL is less selective and thus less desirable than the sophisticated type's last choice (since choices are from X'). This means that the choice does not appear on the ROL of the sophisticated type, thus he can deviate and replace his last choice with the neglectful type's last choice. But because both agents choose their last school conditional on rejection by all higher ranked schools (Shorrer 2019, Lemma 2), the sophisticated type's beliefs are MLRP-lower and this is a contradiction to Proposition 2 of Shorrer (2019), which states that sophisticated agents with higher beliefs apply more aggressively (as the lack of sophistication does not play a role in ROLs of size 1 – except, of course, for the effect of false beliefs from conditioning on previous rejections).

Next, note that the neglectful type's subjective optimal ROL must be identical to his optimal ROL in the original problem (the outside option and the constraint on the size of the ROL were chosen to mimic a situation where his strategy space was restricted to include certain options which appear on his subjective optimal ROL anyway and to not include certain school that did not appear on his subjective optimal ROL anyway).

Lastly, note that if we remove the sophisticated type's access to the outside option, his ROL becomes less aggressive (Shorrer 2019, Theorem 3). And the least selective school on his ROL of size k in this problem is even less selective than the least selective school on his optimal ROL of size $k-m$ (Shorrer 2019, Theorem 1). But the optimal ROL of size k from X' coincides with the optimal ROL of size k from X (Shorrer 2019, Lemma 1). Together with the fact that the least selective school on $r_n(k)$ belongs to X', this completes the proof. $\square$

B. Supplemental Tables Referenced in Text

Table A1. Demographic Information from Primary Experiment.

	Mean	Standard Deviation	25^{th} Pctile	Median	75^{th} Pctile
Female	62.4%				
Graduated High-School in the USA	90.3%				
Degree Involves Math	75.2%				
Age	21.2	3.0	20.0	21.0	22.0
High-School GPA	3.65	0.70	3.60	3.86	4.0
College GPA	3.45	0.37	3.25	3.50	3.70
Nb. Correct Raven's Matrices	5.0	1.2	4.0	5.0	6.0
Correlation-Neglect Measure	0.75	0.51	0.53	0.89	1.03

TABLE A2. Choices in the Correlated and Uncorrelated Settings by Order of Module.

Scenario		Uncorrelated Module First				Correlated Module First				Test of Equality
		$(A \succ B)$	$(A \succ C)$	$(B \succ C)$	Other	$(A \succ B)$	$(A \succ C)$	$(B \succ C)$	Other	p-value
1.	C: (50, 45, 0)	45.9	48.2	4.7	1.2	51.3	41.3	3.8	3.8	0.66
	U: (50, 90, 0)	8.2	87.1	0.0	4.7	13.8	81.3	0	5.0	0.65
2.	C: (50, 45, 10)	52.9	42.4	3.5	1.2	47.5	46.3	2.5	3.8	0.80
	U: (50, 90, 20)	9.4	90.6	0.0	0.0	11.3	85.0	0.0	3.8	0.22
3.	C: (50, 20, 0)	72.9	21.1	5.9	0.0	76.3	16.3	6.3	1.3	0.71
	U: (50, 40, 0)	52.9	38.8	5.9	2.4	46.3	42.5	6.3	5.0	0.89
4.	C: (50, 20, 10)	78.8	12.9	7.1	1.2	85.0	12.5	2.5	0.0	0.44
	U: (50, 40, 20)	75.3	20.0	1.2	3.5	60.0	30.0	6.3	3.8	0.02^{**}
5.	C: (50, 55, 0)	21.2	76.5	1.2	1.2	32.5	61.3	1.3	5.0	0.14
	U: (50, 100, 0)	9.4	87.1	1.2	2.4	6.3	88.8	0.0	5.0	0.52
6.	C: (75, 60, 0)	20.0	50.6	25.9	3.5	30.0	40.0	20.0	10.0	0.20
	U: (75, 80, 0)	12.9	77.7	1.2	8.2	12.5	75.0	6.3	6.3	0.24
7.	C: (75, 60, 30)	28.2	42.4	22.4	3.7	32.5	33.8	28.8	5.0	0.76
	U: (75, 80, 40)	14.1	77.7	1.2	7.2	15.0	75.0	0.0	10.0	0.97
8.	C: (80, 60, 0)	21.2	34.1	38.8	5.9	27.5	23.8	41.3	7.5	0.39
	U: (80, 75, 0)	15.3	55.3	20.0	9.4	13.8	60.0	17.5	8.8	0.95
9.	C: (80, 60, 40)	31.8	25.9	40.0	2.4	28.8	21.3	38.8	11.3	0.19
	U: (80, 75, 50)	25.9	41.2	21.2	11.8	18.8	50.0	21.3	9.5	0.73

Notes: This table summarizes the ROLs chosen in each matched pair of scenarios by which module subjets saw first. All numbers presented (with the exception of the last column) are percentages of responses seen within a module. Columns $(A \succ B)$, $(A \succ C)$, and $(B \succ C)$ present the fractions of subjects reporting each of those ROLs, and column "other" reports the fraction of subjects reporting one of the (dominated) strategies $(B \succ A)$, $(C \succ A)$, or $(C \succ B)$. The final column present the p-value associated with Fisher's exact test for differences across populations who saw the uncorrelated or correlated module, using the full distribution of choices without aggregating dominated ROLs. $^{**}p < 0.05$.

TABLE A3. Choices in the Correlated and Uncorrelated Settings (Using First Module Only).

Scenario		Rank-Order List $(A \succ B)$	$(A \succ C)$	$(B \succ C)$	Other	Test of Equality (p-values) Full Dist.	$(A \succ B)$	$(A \succ C)$
1.	C: (50, 45, 0)	51.3	41.3	3.8	3.8	<0.01***	<0.01***	<0.01***
	U: (50, 90, 0)	8.2	87.1	0.0	4.7			
2.	C: (50, 45, 10)	47.5	46.3	2.5	3.8	<0.01***	<0.01***	<0.01***
	U: (50, 90, 20)	9.4	90.6	0.0	0.0			
3.	C: (50, 20, 0)	76.3	16.3	6.3	1.3	<0.01***	<0.01***	<0.01***
	U: (50, 40, 0)	52.9	38.8	5.9	2.4			
4.	C: (50, 20, 10)	85.0	12.5	2.5	0.0	0.17	0.12	0.19
	U: (50, 40, 20)	75.3	20.0	1.2	3.5			
5.	C: (50, 55, 0)	32.5	61.3	1.3	5.0	<0.01***	<0.01***	<0.01***
	U: (50, 100, 0)	9.4	87.1	1.2	2.4			
6.	C: (75, 60, 0)	30.0	40.0	20.0	10.0	<0.01***	<0.01***	<0.01***
	U: (75, 80, 0)	12.9	77.7	1.2	8.2			
7.	C: (75, 60, 30)	32.5	33.8	28.8	5.0	<0.01***	<0.01***	<0.01***
	U: (75, 80, 40)	14.1	77.7	1.2	7.2			
8.	C: (80, 60, 0)	27.5	23.8	41.3	7.5	<0.01***	0.06*	<0.01***
	U: (80, 75, 0)	15.3	55.3	20.0	9.4			
9.	C: (80, 60, 40)	28.8	21.3	38.8	11.3	0.02**	0.68	<0.01***
	U: (80, 75, 50)	25.9	41.2	21.2	11.8			

Notes: This table summarizes the ROLs chosen in each matched pair of scenarios. All numbers presented (with the exception of the final three columns) are percentages of responses seen within a module using the first module of each treatment only. Columns $(A \succ B)$, $(A \succ C)$, and $(B \succ C)$ present the fractions of subjects reporting each of those ROLs, and column "other" reports the fraction of subjects reporting one of the (dominated) strategies $(B \succ A)$, $(C \succ A)$, or $(C \succ B)$. The final 3 columns present p-values associated with tests for differences across the correlated and uncorrelated presentations. The column marked "Full Dist." presents the results of Fisher's exact tests of differences in the distribution of the six possible ROLs by correlation condition. The following two columns present two-sample difference-of-proportions tests, comparing the proportion picking each of the focal strategies across correlation conditions. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

TABLE A4. Predicting Correlation-Neglectful Preference Reversals.

	(1)	(2)
Enke-Zimmermann Measure	0.323 (0.120)***	0.263 (0.125)**
EZ Missing	0.212 (0.100)**	0.171 (0.105)
Raven's Matrices Performance		-0.046 (0.025)*
Female		0.036 (0.060)
High School GPA		0.007 (0.042)
College GPA		-0.090 (0.082)
Attended High School in USA		-0.043 (0.099)
Math		0.052 (0.073)
Constant	0.220 (0.090)**	0.751 (0.331)
# of observations	157	157
R^2	0.045	0.080

Notes: This table presents OLS regressions of our measure of the rate of correlation-neglectful preference reversals on the Enke-Zimmermann measure of correlation neglect. The Enke-Zimmermann measure is treated as missing if it is measured outside of the unit interval, in which case the variable "EZ Missing" is set to 1. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

TABLE A5. Assessing the Rate of Aggressive Strategies Within Module 1

	Should Be Aggressive	Actually Aggressive	
Correlated Arm	-0.212***	0.001	0.012
	(0.023)	(0.035)	(0.039)
Constant	0.742***	0.538***	
	(0.016)	(0.026)	
Scenario FE	No	No	Yes
Observations	1650	1650	1650

Notes: This table reproduced the results of Table 5 while restricting the data to only the first module presented to subjects. Standard errors, clustered by subject, are presented in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C. TIME TRENDS

Figures A1 and A2 both show the fractions of aggressive and diversified portfolio choices, the former in the correlated arm of the experiment, the latter in the uncorrelated arm. Recall that the order in which subjects faced each scenario (see Table 1) was randomly determined at the subject level. This allows us to examine if the tendency towards aggressive or diversified portfolios changed as experience answering these questions accumulated.

To illustrate the interpretation of these figures: the bars presented in time period 1 of Figure A1 show that, averaging over all scenarios presented first in the correlated treatment arm, subjects chose the aggressive portfolio, 39% of the time and the diversified portfolio 36% of the time. The fourth scenario subjects faced in the correlated treatment arm resulted in the aggressive portfolio being chosen 40% of the time on average, while the diversified portfolio is chosen 41% of the time.

Both graphs show that there is little evidence of time trends in either condition, as the fraction of aggressive and diversified portfolios remain largely stable over the course of the experiment.

FIGURE A1. Time Trends in Portfolios Choices: Correlated Treatment Arm.

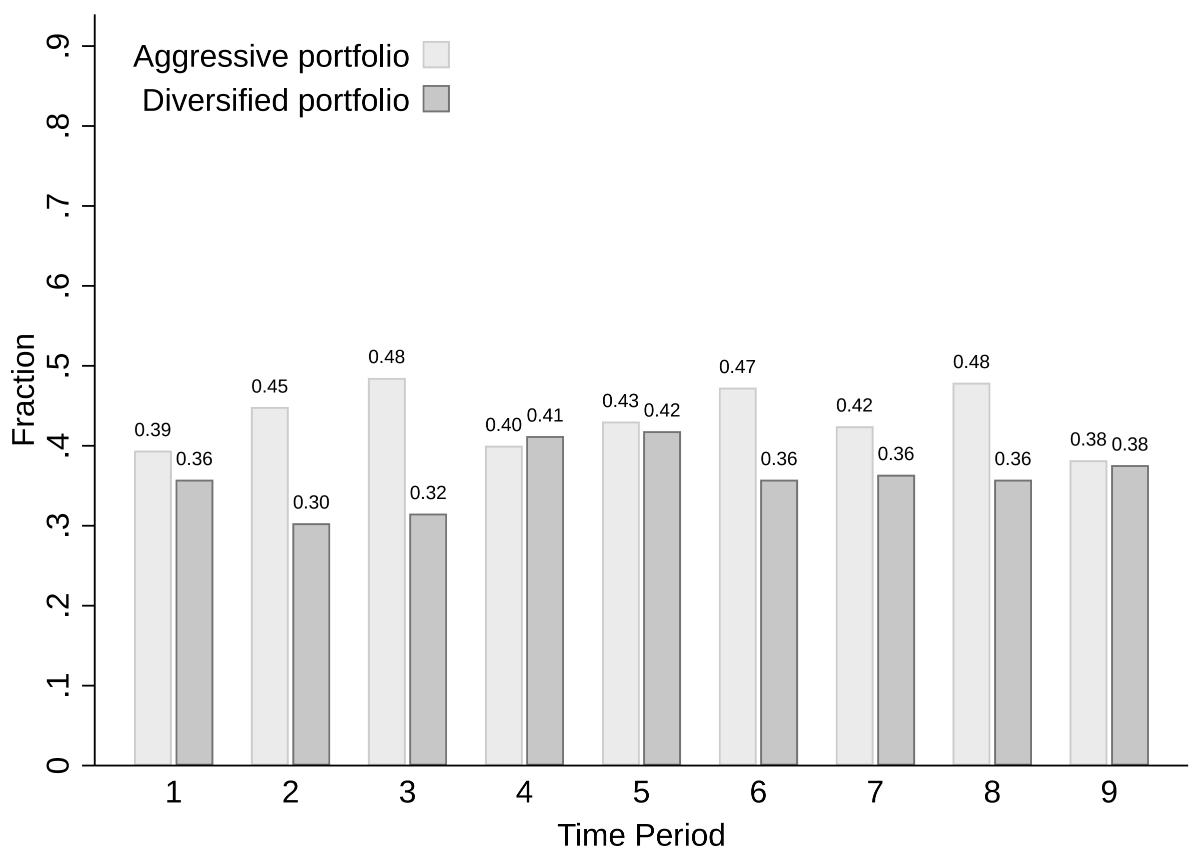
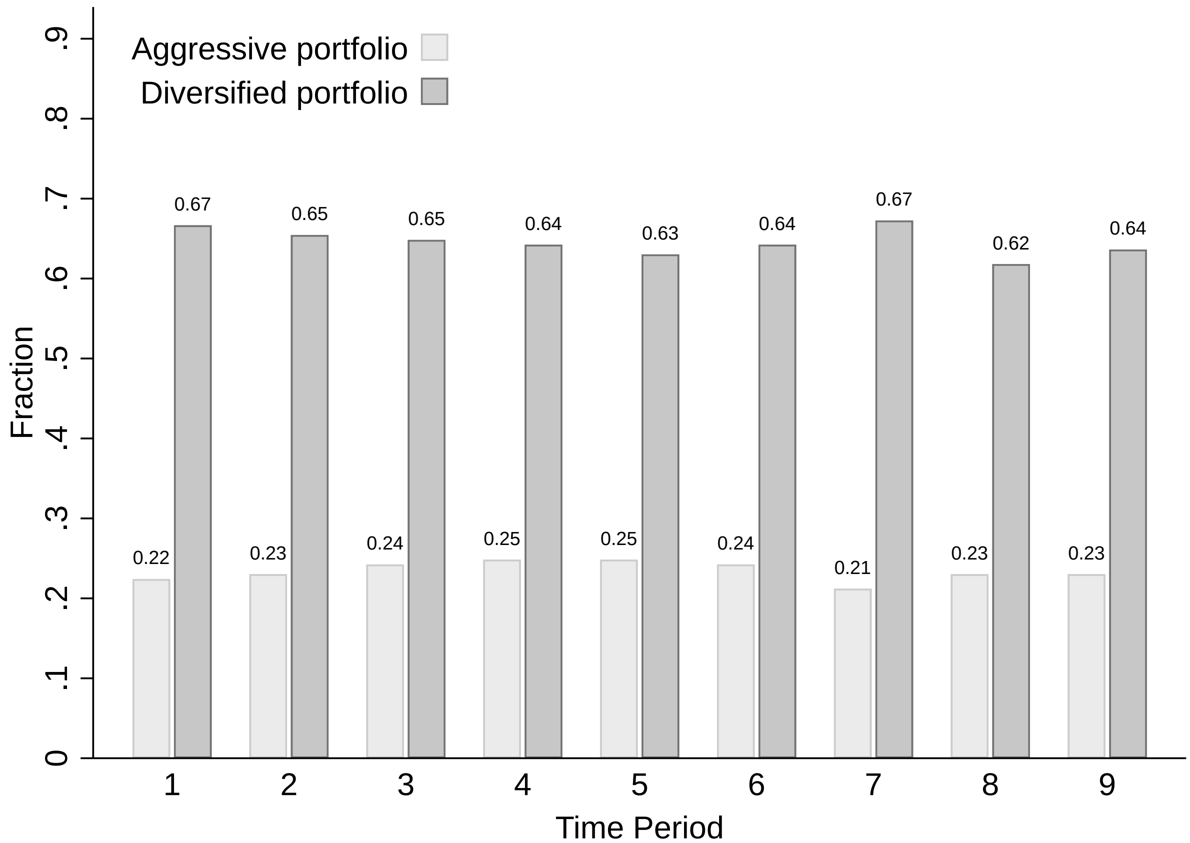


FIGURE A2. Time Trends in Portfolios Choices: Uncorrelated Treatment Arm.



D. Alternative Explanations for Results of the Primary Experiment

In this Appendix we consider several alternative explanations for the results of our primary experiment.

D.1. **Aversion to Schools Dominated as Singleton Applications.** Recall that in our leading example (matching Scenario 1), subjects were more likely to submit an ROL rationally foregoing the middle school when that school had an admissions threshold of 90 on an independent test as opposed to when it had an admissions threshold of 45 on the common test. While both framings result in the same distributions of outcomes when the ROL (best, middle) are applied, note that they would result in different outcomes if middle were listed in isolation. In the uncorrelated framing, the middle option is formally dominated as a singleton application: it has a lower utility and a higher admissions threshold than the best school. In the correlated framing, it is not dominated: while it does have a lower utility, it also has a lower threshold. If subjects are irrationally averse to including such options in a multi-school ROL, this aversion could guide them towards optimal behavior (for reasons different than our focus in this study).

Note, however, that while this concern is present in our leading example, it is not present in scenarios 3, 4, 8, or 9. In all such cases, compared to applications to top school A, applications to middle school B yields a lower utility with a higher chance of admissions regardless of framing. The continued presence of qualitatively large differences in application behavior in these environments alleviates the worry that this potential aversion explains the results we have documented.

D.2. **Models of Choice-Set Dependence.** We consider next the potential explanatory value of a class of choice-set-dependent models of recent prominence in the behavioral economics literature. These models consider how the distribution of attributes in a choice set influence the decision weights placed on those attributes, with greater weight meant to capture the devotion of additional attention. In the focusing model of Köszegi and Szeidl (2012), it is assumed that an attribute with a larger range of values receives more decision weight. In the relative thinking model of Bushong et al. (2020), it is assumed that an attribute with a more narrow range gets more attention. In the salience model of Bordalo et al. (2012), the key predictions come from their assumptions of ordering and diminishing sensitivity, which at times point in the direction of either of the previous models. For a recent paper carefully comparing these theories and their empirical performance in explaining experimental purchasing decisions, see Somerville (2019).

When applying these frameworks to our setting, we believe it is most natural to imagine the student to be considering two attributes: payoffs and admissions thresholds. Payoffs are held constant in our design, but admissions thresholds (and their ranges) differ. Referencing Table 1, note that the manner in which thresholds change makes the range of thresholds in the uncorrelated treatment larger in some scenarios (1, 2, 5-7), smaller in other scenarios (4, 9), and unchanged in yet others (3, 8). The fact that we document apparent neglect of the safety option that is most attractive on the admissions-threshold dimension across all of these variants suggests that choice-set-dependent models based on comparisons of range do not provide a natural explanation for our results.[34]

D.3. **Independence Neglect.** We have interpreted the different preferences that subjects express in the correlated and uncorrelated treatment arms of the primary experiment as evidence of correlation neglect. An alternative conceptual possibility is independence neglect: acting as if outcomes are correlated by default, and neglecting to account for their independence when it is experimentally imposed. Independence neglect could generate some findings that we have documented, but two elements of our results suggest that it is not relevant. First, recall that choices in the uncorrelated treatment are those most closely aligned with choices in our transparent lotteries. While it is perhaps reasonable to suggest that real-life admissions decisions typically are correlated, and thus that the default behavior should assume correlation, it is less plausible to suggest that such a presumption of correlation exists for transparent lotteries. Second, in scenarios involving a second-best school that is unattainable given rejection from the first-best (e.g., scenario 5), independence neglect predicts no difference in behavior across conditions, whereas correlation neglect predicts the observed differences. These considerations, combined with the large psychology literature demonstrating a tendency towards a default assumption of independence (for discussion, see Fiedler and Juslin 2006), lead us to believe that independence neglect is unlikely to meaningfully affect our results in our primary experiment. Furthermore, the results of our structural experiment directly rule out independence neglect in a very similar experimental context.

D.4. **Preferences for Randomization.** We interpret our finding of within-subject preference reversals to be strong evidence of incorrect processing of correlated environments. This interpretation relies on the assumption that behavior would not respond to framing if all elements of the decision environment were fully understood,. However, several recent works have examined cases where subjects hold an explicit preference for randomization (see, e.g., Agranov and Ortoleva 2017, Dwenger et al. 2018, Cerreia-Vioglio et al. 2019); in the presence of such preferences, inconsistent choice need not reflect a true preference reversal.

Four pieces of evidence suggest that preference for randomization have little role in our results. First, note that when subjects faced their first module of school-choice decisions, they did not know that an additional round of equivalent-but-differently-framed scenarios would follow. Without knowledge that two iterations of each question would occur, the underlying motivations that guide intentional randomization would not be triggered. Second, we note that intentional randomization alone would not generate the stark asymmetry observed: while it could predict different answers within-subject, it would not predict the strong tendency for aggressive applications specifically under correlated framing. Third, a preference for randomization would not explain why choices

[34]Note that only the ordering assumption in Bordalo et al. (2012) depends directly on range. Turning to the assumption of diminishing sensitivity, we note that this assumption considers the salience implications of shifting the values of attributes to be larger across both options considered. Since most of our scenarios vary only the threshold for middle program while holding the thresholds for the other programs constant, this assumption does not apply to our setting as written.

made in the absence of correlation were more in line with choices in transparent lotteries. Finally, even if a preference for randomization obfuscates the interpretation of within-subject preference reversals, the between subject contrasts we have presented would remain valid. Given these issues, we believe that preference for randomization does not provide a systematic account of the results we have documented.

E. An Auxiliary Experiment Testing Debiasing Interventions

The results of the main text suggest that correlation neglect can negatively influence students' application strategies. We now turn to a practical question: can these negative effects be mitigated through "debiasing" interventions?

E.1. **Design and Deployment of the Debiasing Experiment.** The debiasing experiment was built to closely mirror the school-choice scenarios used in our primary study. In the control conditions, the scenarios were presented in a manner that closely aligns with our previous correlated and uncorrelated treatment arms. Behavior within these control arms serves as our benchmark for behavior with "unmitigated correlation neglect" or "no correlation neglect," respectively. The treatment conditions consisted of five candidate means of guiding participants to wiser decisions.

While the experimental task in the debiasing experiment is very similar to that in our primary study, two design considerations are notably different.

The first key difference involves the length and complexity of the study. Unlike the primary study, which involved multiple decisions per subject and within-subject manipulations, the debiasing experiment consisted of only a single incentivized scenario and a fully between-subjects design. Subjects faced either one of the two control conditions or one of the five debiasing conditions. We implemented this shorter and simpler design based on our concern that debiasing treatments may be interpreted differently, or may be unusually salient, when they are viewed in immediate contrast with each other. Because subjects only submit a single application when participating in the field applications that motivate our study, we view this single-decision design as a more externally valid way of testing potential interventions.

The second key difference involves sample size and our approach to managing statistical power. Whereas the primary experiment was aimed at establishing the presence of an effect, the debiasing experiment was aimed at measuring potentially small differences in effect sizes. To measure these small differences, substantially greater statistical precision is needed. This led us to pursue a much larger sample size, and thus to move out of the lab and onto an online experimental platform (Prolific). It also led us to gear our statistical analyses towards data with all scenarios pooled rather than focusing attention on differences across scenarios.

Based on these considerations, our experiment followed a simple structure. We begin by describing the control conditions, then describe the debiasing treatments. The complete text of the experiment and a copy of the preregistration (which we follow precisely) are available in the Materials Appendix.

After providing informed consent and completing a CAPTCHA task, subjects were presented with the set-up of the school-choice scenarios. The decision environment remained the same as in the primary study: subjects submit an ROL listing two of three colleges, these colleges would make admissions decisions by comparing randomly drawn test scores to the posted minimum test scores, and the college where the student is admitted determines the bonus. As in the primary study, test scores are drawn from uniform distributions over the integers 0 to 99. The bonuses for enrollment in the three colleges were again \$10, \$5, or \$2.50. In the control condition, we randomized whether we presented a decision environment with a single test score (the correlated condition) or independent test scores at each college (the uncorrelated condition). Because the treatment arms are meant to assess interventions to help subjects correctly think about the impact of correlation, those arms all presented the single-test-score version of these materials.

As in the primary experiment, after reading the initial explanation of the school-choice simulation, all subjects participated in a practice round. In this practice round, subjects faced an example set of three colleges, submitted their ROL, then saw the results. The results screen included their simulated test scores, the college where they were admitted, and the bonus they would earn if this were the incentivized scenario.

Having completed these introductory materials, subjects were then presented with a single incentivized scenario that is the focus of this experiment. Subjects were randomly presented with scenario 1, 2, or 5 from Table 1. After submitting their ROL, final results were simulated. A results screen presented their simulated test scores, the college where they were admitted, and the bonus earned. This concluded the experiment.

The five treatment arms of our experiment each involved additions to the template above that could conceivably help guide respondents to wiser choices. We describe each arm below.

- **Lottery:** The lottery treatment was designed to confront subjects with the probabilities of different outcomes that would arise based on their submitted ROL. After subjects enter their ROL in the incentivized scenario, they see a screen with the header "Confirming Your Answer." This screen reminds the subject of their first and second choice, then reports resulting probabilities of enrolling in either choice as well as the probability of being unmatched. The subject then has the opportunity to either confirm or modify their decision. If the decision is modified, the process repeats (up to a maximum of six attempts).[35]
- **Score Explanation:** The score explanation treatment was designed to help subjects understand how different ranges of test scores mapped to different outcomes. After the subject completed the practice round, they faced a screen with the header "Additional Explanation." On this screen, another example scenario was presented. After presenting the table communicating the bonuses and minimum tests scores, the subject saw a series of bullet points explaining where they could match based on their draw of test scores. After this screen the subject continued to the incentivized scenario.

[35] This treatment arm was motivated by pop-up screens used to combat deviations from truthful preference reporting in the Israeli Psychology Masters Match (Hassidim et al. 2017b, 2021), and later in the American Genetic Counseling Admissions Match (Peranson 2019). Further note that the limit of six attempts was only hit by one subject.

- **Sequential:** The sequential treatment was designed to examine a means of eliciting ROLs that potentially makes the necessary contingent reasoning more salient. Rather than having respondents submit a full ROL, preferences are instead elicited sequentially. In both the training materials and the incentivized scenario, subjects are first asked to submit their first choice, and then on a separate screen are reminded of their first choice and asked to submit their second choice.
- **Extra Practice:** The extra practice treatment presents subjects with an extended practice module prior to the incentivized school-choice scenario. While all subjects complete one practice round, subjects in the extra practice treatment go on to an additional page that presents a new example scenario. Rather than submitting an ROL in this case, subjects must answer two questions. These questions posit that a student submits either the aggressive or diversified strategy, and the subject must indicate where the student will match given that submission and their reported test score. Subjects who answer incorrectly are told so, and must try again until they correctly answer the questions (up to a maximum of 16 attempts).[36]
- **UK Intervention:** The UK intervention treatment presents subjects with text closely mirroring guidance that has been provided to students in the UK's UCAS system (described in Section 1.1). Immediately prior to facing the incentivized scenario, subjects are told in bold font "When deciding on where to apply, keep in mind that your second choice should be used as a backup. You will ONLY enroll in your second choice if your exam score is too low to be admitted to your first choice AND your exam score is high enough to be admitted to your second choice."

Because a central goal of our experiment is comparing each arm to the control baselines, we assigned subjects to treatment in a manner that resulted in equal probability of appearing in each treatment arm but which oversampled the control arms. Each treatment arm was assigned with probability $\frac{1}{8}$ (targeting approximately 250 observations per arm) and each control arm was assigned with probability $\frac{3}{16}$ (targeting approximately 375 observations per arm).

The debiasing experiment was deployed on the Prolific experimental platform in December, 2021. On average, participation in the experiment took 5 minutes. All subjects received a $1 fixed payment for participation and earned an average bonus of $5.39 from the incentivized school-choice scenario. 1,999 subjects completed the experiment, with an average age of 34 years and a 49-51 male/female gender distribution.

E.2. **Debiasing Results.** Appendix Table A6 presents the distribution of ROLs submitted in each treatment arm of the experiment.[37] Focusing attention first on the first two columns, which describe behavior in the control arms, we find that we replicate the main effect of the primary experiment. In the uncorrelated control treatment, 20.6% of subjects pursue the aggressive strategy and 63.0% of subjects pursue the diversified strategy. In the correlated control treatment, 37.2% of subjects

[36]The limit of 16 attempts was only hit by one subject.

[37]For scenario-specific distributions, see Appendix Table A7.

Table A6. Distribution of ROLs in Debiasing Experiment.

	Uncorrelated Control	Correlated Control	Lottery	Score Explanation	Sequential	Extra Practice	UK Intervention	Total
$A \succ B$ (Aggressive)	20.6	37.2	37.6	35.6	35.2	33.6	28.0	32.2
$A \succ C$ (Diversified)	63.0	38.7	47.3	40.6	46.9	48.0	52.8	48.4
$B \succ A$	7.1	8.6	4.3	7.5	7.0	3.1	6.7	6.5
$B \succ C$	2.8	11.8	7.8	11.3	7.8	10.5	9.1	8.6
$C \succ A$	4.8	1.8	0.4	3.8	2.7	3.1	1.2	2.6
$C \succ B$	1.7	1.8	2.7	1.3	0.4	1.6	2.4	1.7

Notes: This table presents the percentage of respondents submitting each possible rank order list (ROL) in each treatment arm of the debiasing experiment. The final column presents the distribution of ROLs aggregating across all treatment arms. $n = 1,999$.

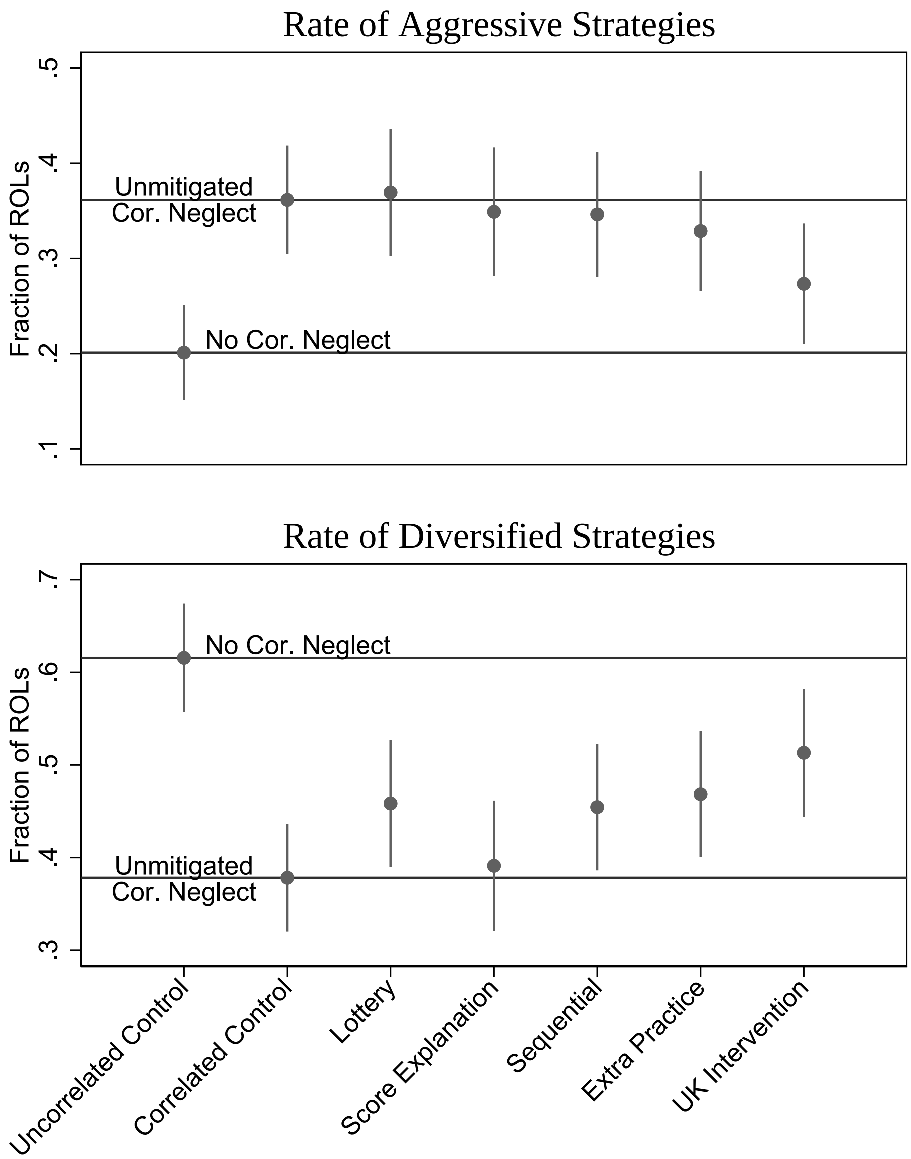
pursue the aggressive strategy and 38.7% of subjects pursue the diversified strategy. Independent of our interest in debiasing interventions, this pattern of behavior in the control arm is reassuring: it demonstrates that the main results of the primary experiment are reproducible in a different and much larger sample.

Turning next to the remainder of columns, we see that the rate of the aggressive ROL in the debiasing conditions ranges from 28.0% to 37.6%. Several debiasing treatments have rates of the aggressive ROL in close vicinity to the rate in the correlated control arm, and all debiasing treatments have rates of the aggressive ROL at least 7.2 percentage points higher than occurs in the uncorrelated control arm. Similarly, the rates of the diversified ROL are at times in close vicinity to the rate in the correlated control arm, and all debiasing treatments have rates of the diversified ROL at least 10.2 percentage points higher than occurs in the uncorrelated control arm. Taken together, while some variation exists across treatments, these results indicate that the debiasing interventions as a group lead to only modest improvements to decision-making.

To formally statistically assess these differences, we conduct OLS regressions of the form $\sigma_i = \sum_t \beta_t \times I(Treatment_i = t) + \mu_s + \epsilon_i$. These regressions predict the strategy σ pursued by person i with a treatment-arm (t) specific rate β_t. The term $I(Treatment_i = t)$ equals one if subject i was assigned to treatment t and is zero otherwise. The term μ_s denotes scenario-specific fixed effects and ϵ_i is the usual error term. The results from using this framework to predict aggressive or diversified strategies are reported in Appendix Table A8 and are visually presented in Appendix Figure A3.

Directing attention first to the top panel of Appendix Figure A3, note that the two estimates presented on the left indicate the rate of aggressive strategies pursued in the uncorrelated and correlated control arms. These serve as our baselines for behavior that is not influenced by correlation neglect or behavior that is influenced by unmitigated correlation neglect. The five estimates on the right present the rate of aggressive strategies in each treatment arm (sorted by effect size). As is visually apparent, the only treatment with an economically or statistically significant impact was the UK intervention. Relative to the unmitigated correlation neglect baseline, the UK intervention reduced the rate of the aggressive strategy by 8.8 percentage points ($s.e. = 3.7$, $p = 0.019$), or

FIGURE A3. Strategies in the Debiasing Experiment.



Notes: This figure presents estimates from OLS regressions of indicators for pursuing either the aggressive strategy (top panel) or the diversified panel (bottom panel) on indicators for each treatment arm and each scenario. Dots indicate the point estimates and the associated lines indicate 95%-confidence intervals (based on heteroskedasticity-robust standard errors). The regressions are formally reported in the first two columns Appendix Table A8. As in that table, the plotted estimates treat Scenario 1 as the baseline.

54.9% ($s.e. = 20.3\%$) of the difference attributed to correlation between control conditions. Examining the second panel, we again see that the UK intervention resulted in the largest improvement in decision-making. Relative to the unmitigated correlation neglect baseline, the UK intervention increased the rate of the diversified strategy by 13.5 percentage points ($s.e. = 4.0$, $p = 0.001$), or 56.8% ($s.e. = 15.2\%$) of the difference attributed to correlation between control conditions.

Across the other debiasing interventions, we find statistically stronger improvements than were seen for aggressive strategies,[38] although the quantitative impacts remain modest.[39]

These findings persist across a variety of regression specifications of potential interest. Similar results arise in regression analyses that do not include scenario fixed effects, and in regressions that include data only for subjects listing the aggressive or diversified strategies or only strategies in alphabetical order. See Appendix Table A8 for these results.

E.3. **Summary and Interpretation.** Taken as a whole, the results of the debiasing experiment indicate that the mistakes that we attribute to correlation neglect are not simply corrected. We view it as encouraging that an intervention used in the field is at least partially successful in debiasing, and as a practical consideration we recommend that markets facing these issues use text like that of the UK's UCAS to help guide their participants to better decisions. Despite this encouraging finding, the fact that significantly worse decision making occurs in the presence of correlation for any debiasing treatment suggests that correlation neglect cannot easily be eliminated. Even with the best debiasing treatments that we have tested, market organizers should still expect correlation neglect to impact their participants.

While less useful to practitioners, our results concerning the less successful debiasing interventions remain valuable in assessing the overall results of this paper.

The fact that the score explanation and extra practice interventions had minimal effect suggests that the patterns of results we have documented are not easily attributable to inadequately described scenarios. Indeed, in the extra practice treatment, 91% of subjects correctly answered the two additional practice questions on their first attempt. While a reasonable reader may have wondered whether participants were unsure of the nature of the simulation, and thus believed that further explanation would help remove bias, these results assuage such worries.

The fact that the lottery intervention had minimal effects was more surprising to us. However, resistance to debiasing through this means can be rationalized by a hesitance to revise decisions due to either time costs or a variety of psychological biases (such as overconfidence in initial intuitions or default effects). When confronted with the opportunity to revise their ROL submission after seeing the probabilistic outcomes it generates, 89% of subjects continued with their original submission. Among the few subjects who did revise their decision, the most common revision was changing an aggressive ROL to a diversified ROL. This is consistent with this intervention having the desired effect among respondents willing to reconsider, but with subjects being willing to reconsider too rarely.

[38] At the 5% alpha-level, the rates of diversified strategies are statistically distinguishable from the unmitigated correlation neglect baseline for the lottery treatment ($p = 0.042$) and the extra practice treatment ($p = 0.023$), and are nearly so for the sequential treatment ($p = 0.055$).

[39] The stronger results for increases in the diversified strategy than for decreases in the aggressive strategy can be partially explained by the debiasing interventions also improving the decision-making of respondents who chose non-focal strategies. In regressions like those conducted for the aggressive and diversified strategies, we find that the lottery, sequential, and extra practice treatments all led to reductions in the rate of pursuing any other strategy that are statistically detectable at the 10% alpha-level.

Finally, the minimal impact of the sequential treatment indicates that our findings are robust to simple interventions meant to subtly guide subjects towards the right contingent reasoning. Note, however, that we do not interpret this finding to mean that helping subjects think about contingent reasoning is categorically not useful—indeed, the success of the UK intervention could potentially be attributed to the fact that it saliently describes the nature of the exact contingency when the second choice is considered.

When considering the success of the UK intervention relative to all other treatments, a notable difference emerges: while all other treatments provide means of guiding the subject to better thinking about correlation, the UK intervention simply tells the subject what to do. In the context of a reasonably complex decision problem, expecting small interventions to assist in understanding of fundamentals may be asking too much. Instead, telling subjects the nature of their optimal strategy in a relatively heavy-handed way may be most useful. However, while doing so was useful in this experiment, it did not fully resolve the problem.

E.4. **Auxiliary Debiasing Experiment Tables.** Auxiliary tables referenced in the debiasing experiment analysis are produced below.

TABLE A7. Comparing ROLs by Scenario and Experiment.

			Rank-Order List			
	Scenario	Experiment	$(A \succ B)$	$(A \succ C)$	$(B \succ C)$	Other
1.	C: (50, 45, 0)	Primary	48.5	44.9	4.2	2.4
		Debiasing	38.1	37.3	12.7	11.9
	U: (50, 90, 0)	Primary	10.9	84.2	0.0	4.9
		Debiasing	16.9	62.9	2.4	17.7
2.	C: (50, 45, 10)	Primary	50.3	44.2	3.0	2.4
		Debiasing	43.0	31.0	13.4	12.7
	U: (50, 90, 20)	Primary	10.3	87.9	0.0	1.8
		Debiasing	25.9	57.1	2.7	14.3
5.	C: (50, 55, 0)	Primary	26.7	69.1	1.2	3.0
		Debiasing	29.5	49.2	9.0	12.3
	U: (50, 100, 0)	Primary	7.9	87.9	0.6	3.6
		Debiasing	19.5	68.6	3.4	8.5

Notes: This table summarizes the distribution of ROLs submitted for the three scenarios included in the debiasing experiment. The distribution of ROLs submitted in same scenario of the primary experiment is included for reference. All numbers presented percentages of subjects in the scenario/experiment combination that submitted the ROL indicated in the header.

TABLE A8. Assessing the Impact of Debiasing Treatments on Focal Strategies.

	Full Sample				Standard Strategists		Alphabetical Strategists	
	Aggressive	Diversified	Aggressive	Diversified	Aggressive	Diversified	Aggressive	Diversified
Correlated Control	0.16***	-0.24***	0.17***	-0.24***	0.24***	-0.24***	0.18***	-0.28***
	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
UK Intervention	0.07**	-0.10**	0.07**	-0.10**	0.10**	-0.10**	0.07*	-0.14***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Lottery	0.17***	-0.16***	0.17***	-0.16***	0.19***	-0.19***	0.16***	-0.22***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Sequential	0.15***	-0.16***	0.15***	-0.16***	0.18***	-0.18***	0.15***	-0.20***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Extra Practice	0.13***	-0.15***	0.13***	-0.15***	0.16***	-0.16***	0.12***	-0.20***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Score Explanation	0.15***	-0.22***	0.15***	-0.22***	0.22***	-0.22***	0.17***	-0.26***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Constant	0.20***	0.62***	0.21***	0.63***	0.25***	0.75***	0.23***	0.71***
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Scenario Indicators	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Observations	1999	1999	1999	1999	1611	1611	1783	1783

Notes: This table presents results from OLS regressions predicting the submission of either the aggressive or diversified ROL (indicated in the header) with dummy variables for the different treatment arms in the debiasing experiment. Heteroskedasticity-robust standard errors are reported in parentheses. The uncorrelated control is the omitted treatment category and serves as the baseline. Scenario 1 is the baseline scenario when scenario indicators are present. The first two columns present the results used to generate Appendix Figure A3. The next two columns reproduce those results, differing only in their exclusion of scenario fixed effects. The next two columns reproduce the primary regressions while restricting the sample to "standard strategists." This excludes all subjects who submitted any ROL other than the the aggressive or diversified strategy. The final two columns reproduce the primary regressions while restricting the sample to "alphabetical strategists". This excludes all subjects who submitted any ROL not in alphabetical order—i.e., $C \succ A$, $C \succ B$, or $B \succ A$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.