

ARTICLE

Defaults are not a panacea: distinguishing between default effects on choices and on outcomes

David A. Kalkstein^{1*} , Fabiana De Lima², Shannon T. Brady³ , Christopher S. Rozek⁴ , Eric J. Johnson⁵  and Gregory M. Walton¹ 

¹Department of Psychology, Stanford University, Stanford, CA, USA, ²Department of Psychology, The Ohio State University, Columbus, OH, USA, ³Department of Psychology, Wake Forest University, Winston-Salem, NC, USA, ⁴Department of Education, Washington University in St. Louis, St Louis, MO, USA and ⁵Columbia Business School, Columbia University, New York, NY, USA

*Corresponding author: David A. Kalkstein, email: dave.kalkstein@gmail.com

(Received 29 June 2021; revised 28 June 2022; accepted 6 July 2022)

Abstract

Recently, defaults have become celebrated as a low-cost and easy-to-implement nudge for promoting positive outcomes, both at an individual and societal level. In the present research, we conducted a large-scale field experiment ($N = 32,508$) in an educational context to test the effectiveness of a default intervention in promoting participation in a potentially beneficial achievement test. We found that a default manipulation increased the rate at which high school students registered to take the test but failed to produce a significant change in students' actual rate of test-taking. These results join past literature documenting robust effects of default framings on initial choice but marked variability in the extent to which those choices ultimately translate to real-world outcomes. We suggest that this variability is attributable to differences in *choice-to-outcome pathways* – the extent to which the initial choice is causally determinative of the outcome.

Keywords: defaults; nudges; education; choice architecture; field intervention

Introduction

Over the past few decades, *defaults* have featured prominently in research, policy and organizations as a go-to “nudge” for encouraging desirable behavior (e.g., Madrian & Shea, 2001; Johnson & Goldstein, 2003; Goldstein *et al.*, 2008; Thaler & Sunstein, 2008; Sunstein & Reisch, 2014). Within a choice set, the default is the preselected option that is assumed if a person takes no alternative action. A robust literature has shown that a given course of action is more likely to be taken when it is positioned as the default (see Jachimowicz *et al.*, 2019; Mertens *et al.*, 2022). As a result, defaults have become celebrated as an easy-to-implement and low-cost intervention

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

with the potential to produce dramatic benefits on outcomes ranging from increasing enrollment in retirement savings plans to willingness to serve as an organ donor (Benartzi *et al.*, 2017).

However, while defaults have robust effects on the choices people make, how those choices translate to ultimate outcomes is not always direct or determinative. Choices often unfold over time and within larger social and institutional contexts. Thus, a crucial next step for research on defaults is to explore the *choice-to-outcome pathway* to understand how defaults at a single decision point interact with the broader context to affect ultimate outcomes. Relatedly, past research demonstrates considerable variability in the size of default effects across studies and contexts (see Jachimowicz *et al.*, 2019; Mertens *et al.*, 2022). As such, it is also important for researchers to further extend the study of defaults into novel domains and explore contextual factors that may explain some of this variability.

In this article, we examine defaults in a large-scale field experiment in an educational setting. We test whether a default intervention among high school students enrolled in Advanced Placement (AP) classes can increase the likelihood that students take a relevant AP exam. Our aims with this research are (1) to test the effectiveness of defaults in a novel large-scale applied context (more than 32,000 students) and (2) to explore how effects from a single default framing unfold over time in a complex institutional setting. More generally, this research joins a growing body of literature seeking to understand the broader contextual factors that allow psychological interventions to take hold and produce meaningful change in real-world outcomes or not (see Walton & Yeager, 2020; Goyer *et al.*, 2021).

Default effects on choices vs outcomes

Seminal research has shown that defaults can have striking effects on people's choices. Madrian and Shea (2001) found that changing the policy at a large company to make automatic enrollment in a 401(k) retirement plan the default boosted employee participation from 37% to 86%. Johnson and Goldstein (2003) found that European countries that set organ donation as the default (i.e., citizens had to take action to opt-out) had consent rates ranging from 86% to 99%, while countries with a default of non-donation (i.e., citizens had to opt-in to donate) had consent rates ranging from 4% to 28%. These studies and others are well-known for demonstrating the potential for defaults to exert outsized influence on consequential choices.

While defaults can produce large effects on immediate choices, the literature also reveals marked variability in the extent to which those choices translate to ultimate outcomes. In Madrian and Shea's (2001) study of enrollment in 401(k)-retirement plans, the decision made at the point of the default manipulation was determinative of the outcome and thus had a large impact on actual savings behavior. But this is not always the case and, for many reasons, the strength of the default effect on eventual behavior can and does vary across applications. For example, while Johnson and Goldstein (2003) found that opt-out countries had organ donation *consent rates* that ranged from 3 to 25 times higher than opt-in countries, actual donation rates, reported in the same paper, were only 1.16 to 1.56 times higher.

Why might choices made at the point of the default be nondeterminative of the ultimate outcome? Often, the choice itself is separated in time and causality from its corresponding outcome. When this is the case, outside agents or institutions can assume intermediary roles in translating individual choices to outcomes (see Willis, 2013). For instance, institutions may formally or informally revisit the decision by representing the choice either to the original decision-maker or to others with decision-making power. Critically, if the choice is represented without a default framing, or without reference to the decision made when the default framing was initially encountered, the original effect of the default may disappear. Moreover, in complex institutional settings, the framing of an initial decision is only one source of influence on ultimate outcomes (Steffel *et al.*, 2019). Intermediaries may have their own incentives for influencing behavior in one direction or another (e.g., Willis, 2013), and social forces such as institutional policies or norms may alter or overwhelm the effect of a default framing before it reaches fruition in ultimate outcomes. For example, in the case of organ donation, hospitals may have policies that would prevent donation regardless of a deceased person's registration status, such as prohibiting donations of diseased organs or honoring family members' objections (Johnson & Goldstein, 2003; Steffel *et al.*, 2019).

Thus, to understand variability in how default framings affect real-world outcomes, we must understand both the structure of the choice-to-outcome pathway and the broader social and institutional forces that affect the final outcome.

Current research: defaults in education

Although educational institutions are critical to children's learning and development, relatively little research has examined defaults in educational contexts. Of relevance to the present research question, education also provides a context in which decisions – such as whether to take achievement tests whose passage could confer important benefits – often unfold over time and are subject to institutional forces such as school-wide policies and norms.

We tested whether defaults implemented in the College Board's Advanced Placement (AP) program could increase student participation in AP exams. The AP program offers high school students across the United States access to college-level curricula in a variety of subjects with end-of-year standardized exams. The central aim of the AP program is to provide students educational opportunities and enhance student learning through providing rigorous coursework. Students who take end-of-year AP exams can report their exam scores on college applications to demonstrate their academic achievements. Scores above a certain threshold earn students course credit and placement into upper-level courses at many colleges and universities. Thus, AP exams offer students an opportunity to both increase the strength of their college applications and get a head start on college coursework.

For research, the AP program provides an opportunity to study defaults in a large-scale educational context. Within our study, we were able to collect a sizeable sample ($n = 32,508$ students), giving us ample power to detect even modest effects and explore multiple sources of variability. Given the vast reach of the AP program, even small effects would be meaningful (see Kraft, 2020). Nationally, over three

million students take AP classes each year. Thus, even a 1% increase in exam-taking rates would represent an additional 30,000 students gaining the opportunity to strengthen their college applications and earn college credit.

Partnering with College Board, we used the AP program's newly developed online platform to implement a default intervention at the initial point of AP exam registration at the beginning of the academic year. We randomized students to one of two conditions. Some students saw the standard version of the registration question which simply asked if they would like to register for the exam with neither option preselected. Other students – those in the default condition – were presented the registration question in an opt-out format. Here “confirm registration” was preselected and participants had to click an alternative box to opt-out. Given the similarity of this opt-out implementation to that of previous research on which it was modeled (e.g., Johnson & Goldstein, 2003), we predicted that students in the default condition would be more likely to (1) sign up for the exam at the initial decision point and (2) take the exam at the end of the year.

However, as the study was carried out, we became increasingly aware of aspects of the AP program that could interrupt the path from initial decision to exam-taking. First, during the long interval between the initial decision (at the beginning of the school year) and exam-taking (in May), students' exam decisions were revisited several times. Exam decisions were formally revisited twice in the school year when AP coordinators (staff members responsible for organizing and administering the AP programs at their schools) and teachers checked with students (or teachers) to submit and confirm official exam orders and a third time when students ultimately had to decide whether to actually show up and take the exam on the scheduled day. In revisits that involved coordinators and teachers, whether they referenced students' initial selections varied widely across teachers, coordinators and schools.

Second, many teachers, schools and states are subject to institutional incentives, policies and norms that may influence exam-taking above and beyond a default intervention. For example, as intermediaries of students' exam registration decisions, AP coordinators and teachers often have baseline incentives to get students to register for the AP exam. These incentives range from informal incentives to provide students with educational opportunities, to formal incentives such as meeting criteria for professional accountability measures or receiving monetary bonuses for student performance.¹ State and school policies also influence exam registration and exam taking. For example, some states have fee waiver policies that reduce the cost of the AP exam for low-income students, and some schools have policies mandating that students take the exam for each AP course in which they are enrolled. There are also strong and variable norms (a) between AP course subjects and (b) across high schools about the taking of AP exams. For example, it is much more common to take the exam in some subjects (e.g., AP US History) than others (e.g., AP Studio Art 3D).

¹As described in the main text, although there is typically incentive for teachers and coordinators to get their students to take and pass AP tests, the strength and nature of these incentives varies by course, teacher, school, district and state. We do not have measures of any of these variables in the present dataset and so cannot examine them empirically. However, they are all orthogonal to the default manipulation, which was randomized at the individual student level (and thus within teachers, schools, districts and states).

Finally, in addition to these factors, students' beliefs and attitudes toward the exam may change over time as the school year unfolds. For instance, as the exam approaches, students may start to consider pragmatics – like fees and studying time – more concretely, and this may influence the subjective value they place on taking the test (e.g., Trope & Liberman, 2003). Students also no doubt learn at varying rates and develop varying beliefs about their abilities in given AP subjects, their interests and the relevance of a subject to their future goals. Although the present data provided no measures of these student-level variables, these factors give reasons for students to reevaluate their test-taking decisions when presented with the opportunity to do so.

Taken together, the fact that students' initial exam registration decisions were revisited multiple times opened the door for students and intermediaries to act on incentives, policies and norms that may override students' original decisions. More broadly, the indirectness of the choice-to-outcome pathway created the potential for myriad social and institutional factors to overwhelm any initial default effect before it reached fruition in ultimate outcomes. In the study presented below, we explore these aspects of the AP program as they predict students' exam-registration and -taking.

Methods

Participants

In the 2017–2018 academic year, the College Board conducted a pilot launch of a new online platform for the AP program. The online platform was designed to supplement in-class instruction by providing students and teachers access to online resources, such as additional course content and exam registration tools. The pilot included 108 schools from 14 school districts across 4 states: Kentucky, Oklahoma, Tennessee and Texas. The College Board targeted these states to represent two distinct regions of the country (Kentucky and Tennessee for the South; Oklahoma and Texas for the Southwest). Districts within each of these states were targeted to be somewhat representative of national demographics with a slight skew toward districts that included more rural schools, schools with a higher percentage of under-represented minority students in AP, and schools with lower test scores on previous years' AP exams. Once invited by College Board, the decision of whether or not to participate in the pilot launch of the online model was made by districts. Of the 21 invited districts, 7 declined to participate (reasons that districts declined included administrative challenges, staff burdens, privacy concerns and other considerations). All schools within districts that opted to participate were automatically enrolled in the pilot launch of the new online platform for the AP Program.

Participants were 32,508 high school students (39.1% male; 52.5% female and 8.4% other or no response) in these schools who were enrolled in any AP class. Thirty-four percent of participating students identified as an under-represented racial minority (URM) (American Indian or Alaska Native: 0.8%; Native Hawaiian or Other Pacific Islander: 0.1%; Black or African American: 9.6%; Hispanic or Latino [including Spanish origin]: 23.8%) while 49% identified as non-URM (White: 41.2%; Asian: 7.7%) and 16% identified as other, mixed-race, gave no response or had missing data. Forty-one percent were eligible for an AP fee reduction: 35% of the sample received the fee reduction from College Board based on their family income and 6% received

it based on Oklahoma state government policy for students from low-income households. See Table A1 in Appendix for balance checks.

Procedure

Using the College Board platform, schools created online course sections for each AP class offered at their school. These online sections supplemented in-class instruction by giving teachers and students a virtual space to share resources (e.g., learning materials, practice exam questions) and register for year-end AP exams.

At the beginning of the academic year (in August or September), teachers gave students log-in information to join the virtual space for their AP class. The first time students logged in, they were randomly assigned on an individual basis to either the control (opt-in) condition ($N = 16,045$) or the default (opt-out) condition ($N = 16,463$). Once logged in, students were presented a welcome video and some questions about starting AP. Next, they were shown one of the two exam decision pages.

In the control (opt-in) condition, the exam registration question was presented in a traditional format: “Do you wish to register to take the following AP Exam(s) in May 2018?” The options “Register” and “Do not register” appeared on the screen, with neither option preselected.

In the default (opt-out) condition, the exam registration text read, “You are registered to take the following AP Exam(s) in May 2018” and a button indicating “Confirm registration” was preselected for the AP exam(s) corresponding to each AP class(es) in which students were currently enrolled. Students could choose to change the preselected option to the second option (“No, I wish to remove myself from taking this AP Exam”) if they did not want to register for the exam.

Students saw the exam registration question for each course they were taking. If the student was taking multiple courses, all questions were presented on the same page and were presented in the same format (i.e., randomization occurred at the student-level, so students saw either all opt-in or all opt-out exam registration questions). For both conditions, students’ responses were recorded when they submitted the page. If students did not submit the exam registration page the first time they logged in, they were prompted to submit it each subsequent time they logged in. These subsequent prompts followed the same format (i.e., condition-specific framings) as the original prompts.

Students in both conditions were free to modify their choice on the College Board website until November 1st. By November 1st, AP coordinators or AP teachers were required to check-in with students about their online exam registration decisions and officially order exams through College Board. How student registration was confirmed varied by school. In some schools, teachers checked in with students; in others, an AP coordinator checked in with students; in others still, an AP coordinator checked with teachers to confirm exam orders; and finally, in schools that mandated exam-taking, AP coordinators signed everyone up for the exam regardless of students’ original decisions. In all cases, neither coordinators nor teachers knew which initial exam-registration framing students received (default or control) and thus did not frame their check-ins to be consistent with students’ assigned default condition. The system also did not specify how AP coordinators or teachers should check in

with students about their exam registration; nor did it provide a formal reminder of students' initial exam decisions, although this information was available to coordinators and teachers. Anecdotal evidence suggests that there was considerable variability across schools in the extent to which coordinators and teachers referenced students' initial exam decisions when checking in with students, but we have no way of assessing this.

From November 1st through March 1st, students could change their exam registration status by contacting their AP teacher or school AP coordinator. Exams canceled after November 1st incurred a cancellation fee, and exams ordered after November 1st incurred an additional late-exam fee. By March 1st, AP coordinators or teachers submitted final exam orders to the College Board. This order was designed to be largely redundant with the November 1st orders, but also gave AP coordinators and teachers a final opportunity to modify orders before exam day. Of note, neither the November nor March exam order deadlines included a default manipulation for students to respond to.

Independent variables

Default condition was our primary independent variable. We also measured whether students qualified for a fee reduction for their AP exams as an additional predictor of exam taking. In addition, for each student we had variables indicating which AP subject course(s) they were enrolled in and which high school they attended; however, the school identifier provided to us was an arbitrary number assigned by the College Board to anonymize students and schools, so we could not further explore the moderating role of school characteristics. Finally, for the majority of schools in our sample (71%; 29% missing), we also had a variable indicating whether exam taking was mandated by official school policy (27%) or not (73%). These additional variables were included to measure variability in exam registration and exam taking that was attributable to course subject, school and institutional policies.

Dependent measures

Students' exam registration indicated whether students registered for the exam at the point of the intervention (i.e., when they logged into the online AP platform). This variable was recorded for each course in which a student was enrolled. Exam registration was coded such that "1" represented a submitted registration for a particular exam and "0" represented either a submitted declination to register or no response. This variable only indicates students' selection on the webpage where they were presented with the decision. It does not account for whether they later modified that decision.

Student exam taking represented whether students showed up and completed the exam for a given course in May. This variable reflects whether or not College Board recorded an exam score for the student. The variable was coded such that "1" indicated the student took that particular AP exam and "0" indicated that the student was enrolled in a particular AP course but did not take the exam.

While exam registration and exam taking were our primary outcomes, we also had data on students' exam registration status at multiple points through the year.

November registration status represents students' exam registration status as submitted by their AP coordinator or AP teachers on the November 1st ordering deadline. *March registration status* represents students' exam registration status as submitted by their AP coordinator or AP teachers on the March 1st ordering deadline. All data were provided by the College Board.

Results

Exclusions

Due to an error in data recording, exam taking for AP Macroeconomics was not included in the data used in this analysis. As a result, we excluded from the analysis all exam registration decisions and exam taking variables for AP Macroeconomics ($n = 1518$ observations). This left a total of 56,579 observations of exam registration decisions and exam taking across 32,439 students. Descriptive statistics for the number of observations per condition and number of exams registered for and taken are shown in [Table 1](#).

Analytic strategy

Our four dependent measures (student exam registration, November registration status, March registration status and exam taking) were all recorded for each course a student was enrolled in. Thus, exam registration and taking variables were nested within student, and student was nested within school and course subject. To examine the impact of the default framing on each outcome while accounting for clustering within school and course, we conducted multilevel logistic regressions with default condition entered as a fixed effect student-level predictor and school and course subject entered as random intercepts.² We used the “glmer” function within the “lme4” package in R to conduct all analyses (see Bates *et al.*, 2015).

Primary outcomes

Initial exam registration

First, we analyzed whether the default framing affected the likelihood that students registered for their AP course or courses' corresponding exam(s) through the online

²While students could be enrolled in multiple AP classes, most students in our sample (59%) were only enrolled in one AP course. As a result, it was not possible to accurately estimate a model with a random intercept for student. To account for student level non-independence we also ran a multilevel linear regression predicting the percentage of exams a student registered for (out of the total number of AP classes they were enrolled in). In this model, we included default condition as a fixed effect predictor and school as a random intercept (it was not possible to include a random intercept of course since some students were enrolled in multiple classes). This model produced similar results as reported in the main text for initial registration (default effect: $b = 0.023$, $se = 0.003$, $t(32,335.34) 8.11$, $p < 0.001$), and exam taking (default effect: $b = 0.005$, $se = 0.004$, $t(32,335.10) 1.32$, $p = 0.19$). As in the main text, the default effect on November registration status is diminished relative to the effect on initial registration status. However, the effect reaches significance when using the linear model described above ($b = 0.006$, $se = 0.003$, $t(31,559.84) 2.19$, $p = 0.03$). Also as in the main text, the default effect on March registration status is non-significant when using this linear model ($b = 0.005$, $se = 0.003$, $t(31,569.48) 1.68$, $p = 0.09$), though it is trending.

Table 1. Descriptive statistics for exam registration and exam taking.

Condition	# of exam decisions (% of total observations)	# Registered for exam (% w/in each condition)		# Took exam (% w/in each condition)	
		Yes	No	Yes	No
Control	27,983 (49.5%)	24,861 (88.8%)	3122 (11.2%)	22,852 (81.7%)	5131 (18.3%)
Default	28,596 (50.5%)	25,833 (90.3%)	2763 (9.7%)	23,280 (81.4%)	5316 (18.6%)

Descriptive statistics for the number of observations of exam registration and exam taking for each condition. Displayed percentages are raw percentages based on overall observations (Table 2 presents adjusted probabilities that account for clustering of observations within course and school).

Table 2. Adjusted probability of registering for, or taking, an exam at each time point.

Condition	Initial exam decision	November registration status	March registration status	Took exam
Control	90.0%	96.8%	91.1%	77.1%
Default	91.5%*	96.9%	91.3%	76.9%

Adjusted probability of registering for an exam at each time point or taking the exam. Estimates are derived from the multilevel binomial models described in the main text.

*Indicates that the adjusted probability in the default condition is significantly different than the control condition ($p < 0.001$).

AP platform. We found a main effect of default condition ($b = 0.18$, $se = 0.03$, $z = 6.23$, $p < 0.001$). Students were more likely to register for the exam when registration was preselected than when neither option was preselected. As shown in Table 2, the adjusted probability of registration in the default condition was 91.5% vs 90.0% in the control condition. Thus, this effect reflects a 1.5 percentage-point gain and a 15% reduction in unregistered exams among students taking a relevant AP course.

Exam taking

However, there was no effect of the default framing on the ultimate outcome, whether students actually took the exam at the end of the year ($b = -0.01$, $se = 0.02$, $z = -0.51$, $p = 0.61$). The adjusted probability of taking the exam was 76.9% in the default condition vs 77.1% in the control condition (see Table 2).

Exploratory analyses

Next, we sought to explore our data to understand what factors may have interfered with the translation of the initial default effect on students' registration to their ultimate exam taking.

November and March registration

In an effort to understand whether revisiting students' exam decisions multiple times throughout the year may have interfered with the initial default effect, we analyzed whether the default effect emerged at each point when the decision was formally revisited. Importantly, both November and March registration statuses were submitted by AP coordinators or AP Teachers on behalf of students, and decisions were revisited without a default framing manipulation.

First, we tested whether the default effect persisted after AP coordinators and teachers revisited students' exam decisions to submit an official order by November 1st. Although the direction was the same as at exam registration, there was not a significant effect of the default framing during students' initial exam decisions on their November registration status ($b = 0.05$, $se = 0.04$, $z = 1.48$, $p = 0.14$). The adjusted probability of being registered in November was 96.9% in the default condition vs 96.8% in the control condition (see Table 2).

We then examined the March 1st registration status. As with the November registration status, we did not find a significant effect of the initial default framing

on students' March registration status ($b = 0.01$, $se = 0.03$, $z = 0.25$, $p = 0.81$). The adjusted probability of being registered in March was 91.3% in the default condition vs 91.1% in the control condition (see Table 2).

Unsurprisingly, students' registration status predicted actual exam taking more strongly when this status was closer to the exam. The zero-order correlation between exam taking and students' initial exam decisions (when the default framing was encountered) was $r = 0.39$; the correlation between exam taking and November registration status was $r = 0.60$ and the correlation between exam taking and March registration status was $r = 0.80$.

Institutional factors that predict exam taking

Next, we returned to our primary outcome of exam taking and explored the impact of (1) state and College Board policies that reduce exam fees for students and (2) school and course-specific influences on exam taking.

First, to examine whether fee-reduction policies predicted the likelihood of an exam being taken, we ran a multilevel binomial regression with default condition (coded as $-1 =$ control condition; $1 =$ default condition), fee-reduction status (dummy coded so that standard/no fee reduction was the reference group, and receiving state funding and receiving a fee reduction through the College Board each had a variable indicating whether the exam qualified for that type of fee reduction) and their interactions as fixed effect predictors. As before, we also included random intercepts for school and for course subject.

Using this model, we again found no significant effect of the default framing on exam taking ($b = -0.01$, $se = 0.02$, $z = -0.36$, $p = 0.72$). We did find effects of fee reduction policies. The adjusted probability of exam-taking for exams that received fee reduction through state policies was 96%; it was 84% for exams that received fee reduction through the College Board; and 63% for exams that did not receive any fee reduction. Exams were more likely to be taken when students received a fee reduction through state funding than when they received a fee reduction from the College Board ($b = 1.46$, $se = 0.07$, $z = 22.06$, $p < 0.001$). In turn, exams were more likely to be taken when students received fee reductions from the College Board than when they received no fee reduction at all ($b = 1.11$, $se = 0.03$, $z = 32.96$, $p < 0.001$). Fee reduction status did not interact with default condition to predict exam-taking (all interactions of default condition with dummy coded fee reduction variables had $p > 0.75$).³

³A reader may wonder whether there was an effect of policies on students' initial exam decisions. Using the same model described here to predict students' initial exam decisions, we found that exams were more likely to be registered for if the student qualified for state funding to reduce the fee ($b = 1.41$, $se = 0.06$, $z = 22.44$, $p < 0.001$) or for College Board's fee reduction policy based on income ($b = 0.91$, $se = 0.04$, $z = 21.44$, $p < 0.001$) than if the student did not qualify for fee reduction. Moreover, state funding interacted with the default condition. The default framing increased exam registration (adjusted probability of registering: 87.4% vs 84.8%) for students who did not qualify for any fee reduction ($b = 0.11$, $se = 0.02$, $z = 6.10$, $p < 0.001$) but did not have a significant effect (96.2% vs 96.2%) for students who qualified to receive state funding to reduce exam fees ($b = 0.001$, $se = 0.05$, $z = 0.03$, $p = 0.98$). The default effect for students who qualified for the College Board's fee reduction fell in between these two (94.3% vs 93.5%) and was significant ($b = 0.07$, $se = 0.03$, $z = 2.30$, $p = 0.02$) though it was not significantly different from either default effect

To compare the amount of variance explained by fee reduction policies and default framing, we compared the fit of a model that included fee reduction status to one that included only default condition. We found that, including fee reduction status and its interactions as fixed effect predictors significantly improved the fit of the model ($\chi^2(4) 2944.4$, $p < 0.001$, McFadden's Pseudo- $R^2 = 0.06$).⁴ By comparison, including default condition (and its interactions with fee reduction status) did not significantly improve statistical fit from a model that predicted exam-taking likelihood just using fee-reduction status and random effects for school and course ($\chi^2(3) 0.65$, $p = 0.88$, McFadden's Pseudo- $R^2 < 0.001$).

Second, to examine the potential impact of school and course-specific norms, we also used the random effect coefficients of this model to estimate how much variance in the likelihood of taking an exam was attributable to the school and course subject in which a student was enrolled. From these random effect coefficients, we calculated intra-class correlations using the simulation approach described by Goldstein *et al.* (2002).⁵ Using this approach, we estimated that 18.1% of the remaining variance in the likelihood of an exam being taken was attributable to the school a student was attending and an additional 10.0% to the course subject.

Finally, we examined one specific school-level factor that we had the ability to analyze – policies that mandated exam taking (which was available for 71% of the schools in our sample).⁶ To test whether these mandates affected students' actual exam-taking behavior, we ran another multilevel binomial regression with default condition (coded as $-1 =$ control condition; $1 =$ default condition), institutional mandate (coded as $-1 =$ no exam-taking mandate; $1 =$ exam-taking mandate) and their interaction as fixed effect predictors of whether or not a given exam was taken. We again included random intercepts for school and for course subject. We again did not find a significant effect of the default framing on exam taking ($b = -0.01$, $se =$

for students who received no fee reduction nor the effect for students who received state funding ($ps > 0.20$). Using McFadden's Pseudo- R^2 , we estimated that including fee reduction status and its interaction with default condition as fixed effect predictors explained an additional 2.6% of variance in exam registration.

⁴This McFadden's Pseudo- R^2 compares the log-likelihood estimates of a model including fee reduction, default condition and their interaction as a fixed effects and school and course subject as random effects to a model with only default condition as a fixed effect and school and course subject random effects.

⁵This method constructs a simulated distribution of estimated probabilities whose variance is based on a specific random effect from the model. To calculate an ICC, we divided the observed variance of this simulated distribution by an estimate of the total variance in the probability of our outcome (exam registration) derived from the fixed effect estimates of our model.

⁶Given that a school-wide mandate for students to take exams would likely overwhelm the impact of an initial default framing and leave little opportunity to detect its effects on downstream exam taking, we conducted additional analyses on the subset of schools that did not officially mandate exam taking. Of note, this was the majority of schools (73%) for whom we had this information available. In these analyses, we used the same multilevel models reported in the main text to test the effect of the default framing on initial exam registration, November registration, March registration and actual exam taking. These results of these analyses reveal patterns that are highly similar and largely replicate those reported in the main text with the full sample. Specifically, for exams belonging to schools that do not mandate exams, we found a significant effect of default framing on initial exam registration ($b = 0.17$, $se = 0.03$, $z = 5.00$, $p < 0.01$). We did not find significant effects of the default condition on November registration, March Registration or final exam-taking in schools that did not officially mandate exam-taking.

0.02, $z = -0.80$, $p = 0.42$). There was a significant main effect of school policies mandating exam taking ($b = 0.31$, $se = 0.12$, $z = 2.73$, $p < 0.01$). The adjusted probability of exam-taking for exams within schools that mandated exam taking was 85% vs 75% for exams within schools that did not have an official policy mandating exam taking. We did not find a significant interaction between exam-taking mandates and default condition ($b = 0.01$, $se = 0.02$, $z = 0.61$, $p = 0.54$).

Discussion

We found that using a default framing for students' initial registration for an AP exam – presenting exam registration in an opt-out format – significantly increased registration rates compared to a traditional opt-in format. However, the default framing did not significantly affect whether students actually took the exam. Subsequent analysis revealed that the effect of the default framing disappeared the first time students' exam registration status was revisited by their teachers or AP coordinators – which was done without the default framing present. Once the effect disappeared, it never re-emerged. Moreover, while the default effect diminished as soon as students' decision was revisited, the predictive power of students' registration status increased after each revisit; that is, registration status in November was a better predictor of actual exam taking than initial registration choice, while registration status in March was more predictive yet.

Our study highlights the need to consider *choice-to-outcome pathways* when implementing default framings for a given decision. Understanding how decisions unfold over time in complex social and institutional contexts may help policy-makers, institutions and researchers better predict when a default framing will translate to actual outcomes of importance. In cases where an initial decision is determinative of the outcome – as in Madrian and Shea's (2001) enrollment in 401(k) retirement plans – institutions may expect relatively large effects of a default framing on the desired outcome (see also Liebe *et al.*, 2021; Sunstein, 2021). However, when an initial decision is not determinative of the outcome – as in our study – the impact of an initial default framing may depend on if or how that decision is revisited and whether other decision-makers or institutional agents can disrupt the effect. For example, revisiting decisions without the original default framing, and without reference to decisions made under the original default framing, may effectively undo the default effect (see Willis, 2013).

How could the choice-to-outcome pathway be modified to increase the likelihood that a default framing on an initial decision translates to the ultimate outcome? Here, we address this question in the context of our study on AP test-taking as we believe it is illustrative of many problem settings. First, the default framing could have been maintained across subsequent revisits of the initial decision. Second, subsequent revisits could present students' prior decision as a commitment to take the test, as inducing a sense of commitment is a powerful way to get people to follow through (e.g., Moriarty, 1975; Axsom & Cooper, 1985; Koo & Fishbach, 2008; Gollwitzer *et al.*, 2009). Third, the College Board could make the initial decision, in fact, more determinative of the final outcome. That is, it could treat the first decision as the default course of action unless alternative action is taken to change it. Fourth, the College Board could eliminate subsequent revisits altogether, thus making the first decision

– made at the point of the default manipulation – fully determinative. Alternatively, the choice-to-outcome pathway could be designed to be contingent on initial selections such that choices would be revisited only if the original selection was to not register for the exam. Finally, the potential for a default intervention to impact exam taking could have been maximized if it were designed to promote and accompany norms surrounding test-taking (e.g., Davidai *et al.*, 2012).

In weighing these options, it is important to consider not only what will maximize test-taking but when students are best equipped to make this decision – that is, to make the choice that will help them flourish in the long-term, however students define this. Certainly, making a commitment can sustain and organize long-term goal pursuit (Locke & Latham, 1990). However, as the test approaches and students learn more about the subject, their interests and goals, financial costs and opportunity costs in preparing for the test, some make a well-informed decision to not take the test. It is not obvious that students are best served by making a binding commitment at the beginning of the school year, or by making no commitment.

Our research also points to the need to understand the broader social and institutional context when considering the choice-to-outcome pathway and when employing techniques such as defaults to nudge behavior toward positive outcomes (see also Steffel *et al.*, 2019). In our study, the outcome in question – students' exam taking – was impacted more by institutional factors such as fee reduction policies, school-level variables such as exam mandates, and course-specific norms and practices than it was by our default manipulation. In some cases, such institutional factors may have completely overwhelmed any potential impact of a default framing. For example, when schools mandate exam taking for all students enrolled in an AP course, students' initial decisions were overridden when teachers and AP coordinators submitted exam orders to College Board. Thus, similar to how hospital policies that restrict organ donation can reduce the impact of a default framing on actual donation rates, so too can school-wide policies mitigate the impact of a default framing on exam taking.

A future direction for research is to explore how the size of default effects varies with the presence and nature of specific institutional policies and factors. For instance, there should be circumstances in which an initial default framing impacts ultimate outcomes only when it is not overridden by broader institutional policies. We suspect that we did not find interactions in our study between default framing and exam-mandate policies or fee reduction policies on exam taking because the multiple revisits of the initial decision likely disrupted the default effect before it had the potential to interact with institutional policies. However, we expect that such interactions would be more likely to emerge in contexts where the default framing is uninterrupted by multiple revisits of the initial decision.

It is also important to consider the role of material and financial resources as they affect the choice-to-outcome pathway. Presenting students the AP exam registration choice in an opt-out format may influence students' immediate choice, but could fail to increase exam taking for students who simply cannot afford the exam. Indeed, the sizable effect of fee reductions on exam taking points to the importance of relieving financial burdens for AP exams. Overall, our results suggest that the effect of defaults on actual behavior will vary depending on events that happen, and resources that are available, downstream from the decision.

Ultimately, by introducing the concept of a choice-to-outcome pathway, the present research contributes to the literatures on defaults and choice architecture by providing a framework to analyze how single-shot framings translate to eventual behavior. Our findings highlight the need to look beyond a single default framing and consider the full causal and temporal course by which a single choice affects an ultimate outcome. By considering contextual factors – such as if and how a decision is revisited, institutional policies and norms, and financial costs/resources – researchers and practitioners may be better able to understand when a default intervention will be effective in changing real-world outcomes that matter and when it will not. Ideally, this understanding will help institutions design and implement default interventions more effectively and thereby promote better outcomes for individuals and society.

Acknowledgements. We thank our collaborators at The College Board for their support in conducting this study, especially Abby Whitbeck, Scott Spiegel, Rory Lazowski, Robert Garrellick, Lila Schallert-Wygal, Jason Manoharan, Jennifer Merriman and Jennifer Mulhern.

Funding. This work was funded by a grant from the Chan Zuckerberg Initiative and The College Board, which oversees Advanced Placement course curricula and exams. None of the authors have a financial interest in Advanced Placement nor in The College Board.

References

- Axson, D. and J. Cooper (1985), 'Cognitive dissonance and psychotherapy: the role of effort justification in inducing weight loss', *Journal of Experimental Social Psychology*, **21**(2): 149–160.
- Bates, D., M. Maechler, B. Bolker and S. Walker (2015), 'Fitting linear mixed-effects models using lme4', *Journal of Statistical Software*, **67**(1): 1–48.
- Benartzi, S., J. Beshears, K. L. Milkman, C. R. Sunstein, R. H. Thaler, M. Shankar, W. Tucker-Ray, W. J. Condon and S. Galing (2017), 'Should governments invest more in nudging?' *Psychological Science*, **28**(8): 1041–1055.
- Davidai, S., T. Gilovich and L. D. Ross (2012), 'The meaning of default options for potential organ donors', *Proceedings of the National Academy of Sciences of the United States of America*, **109**(38): 15201–15205.
- Goldstein, H., W. Browne and J. Rasbash (2002), 'Partitioning variation in multilevel models', *Understanding Statistics*, **1**(4): 223–231.
- Goldstein, D. G., E. J. Johnson, A. Herrmann and M. Heitmann (2008), 'Nudge your customers toward better choices', *Harvard Business Review*, **86**(12): 99–106.
- Gollwitzer, P. M., P. Sheeran, V. Michalski and A. E. Seifert (2009), 'When intentions go public: does social reality widen the intention-behavior gap?' *Psychology Science*, **20**(5): 612–618.
- Goyer, J. P., G. M. Walton and D. S. Yeager (2021), 'The role of psychological factors and institutional channels in predicting the attainment of postsecondary goals', *Developmental Psychology*, **57**(1): 73–86.
- Jachimowicz, J. M., S. Duncan, E. U. Weber and E. J. Johnson (2019), 'When and why defaults influence decisions: a meta-analysis of default effects', *Behavioural Public Policy*, **3**(2): 159–186.
- Johnson, E. J. and D. Goldstein (2003), 'Do defaults save lives?' *Science*, **302**(5649): 1338–1339.
- Koo, M. and A. Fishbach (2008), 'Dynamics of self-regulation: how (un)accomplished goal actions affect motivation', *Journal of Personality and Social Psychology*, **94**(2): 183–195.
- Kraft, M. A. (2020), 'Interpreting Effect Sizes of Education Interventions', *Educational Researcher*, **49**(4): 241–253. <https://doi.org/10.3102/0013189X20912798>.
- Liebe, U., J. Gewinner and A. Diekmann (2021), 'Large and persistent effects of green energy defaults in the household and business sectors', *Nature Human Behaviour*, **5**(5): 576–585.
- Locke, E. A. and G. P. Latham (1990), *A Theory of Goal Setting & Task Performance*. Upper Saddle River, NJ: Prentice Hall.

- Madrian, B. C. and D. F. Shea (2001), 'The power of suggestion: Inertia in 401 (k) participation and savings behavior', *The Quarterly Journal of Economics*, **116**(4): 1149–1187.
- Mertens, S., M. Herberz, U. J. J. Hahnel and T. Brosch (2022), 'The effectiveness of nudging: a meta-analysis of choice architecture interventions across behavioral domains', *Proceedings of the National Academy of Sciences of the United States of America*, **119**(1): e2107346118.
- Moriarty, T. (1975), 'Crime, commitment, and the responsive bystander: two field experiments', *Journal of Personality and Social Psychology*, **31**(2): 370–376.
- Steffel, M., E. F. Williams and D. Tannenbaum (2019), 'Does changing defaults save lives? Effects of presumed consent organ donation policies', *Behavioral Science & Policy*, **5**(1): 69–88.
- Sunstein, C. R. (2021), 'Green defaults can combat climate change', *Nature Human Behaviour*, **5**(5): 548–549.
- Sunstein, C. R. and L. A. Reisch (2014), 'Automatically green: behavioral economics and environmental protection', *The Harvard Environmental Law Review*, **38**(1): 127–158.
- Thaler, R. H. and C. R. Sunstein (2008), *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.
- Trope, Y. and N. Liberman (2003), 'Temporal construal', *Psychological Review*, **110**(3): 403–421.
- Walton, G. M. and D. S. Yeager (2020), 'Seed and soil: psychological affordances in contexts help to explain where wise interventions succeed or fail', *Current Directions in Psychological Science*, **29**(3): 219–226.
- Willis, L. E. (2013), 'When nudges fail: slippery defaults', *University of Chicago Law Review*, **80**(3): 1155–1229.

Appendix

Table A1. Balance checks of student characteristics by default condition.

	Control	Default
Under-represented minority status		
URM	34.2%	34.3%
Non-URM	48.9%	49.1%
Other, mixed-race, no response or missing	17.0%	16.6%
Gender		
Female	51.9%	53.3%*
Male	39.5%	38.7%
Other or missing	8.6%	8.0%
Fee reduction status		
Fee reduced by college board or state policy	41.0%	41.4%*
None	57.0%	56.9%
Missing	1.9%	1.7%

Note: Rows indicate the percentage of students within each condition of the analyzed sample who belong to each group. Asterisks represent Chi-squared tests of proportional differences between the default and control groups (* $p < 0.05$).

Cite this article: Kalkstein DA, De Lima F, Brady ST, Rozek CS, Johnson EJ, Walton GM (2022). Defaults are not a panacea: distinguishing between default effects on choices and on outcomes. *Behavioural Public Policy* 1–16. <https://doi.org/10.1017/bpp.2022.24>