

Highbrow Films Gather Dust: Time-Inconsistent Preferences and Online DVD Rentals

Katherine L. Milkman

Harvard Business School, Harvard University, Boston, Massachusetts 02163,
kmilkman@fas.harvard.edu

Todd Rogers

Analyst Institute, Arlington, Virginia 22209, trogers@analystinstitute.org

Max H. Bazerman

Harvard Business School, Harvard University, Boston, Massachusetts 02163,
mbazerman@hbs.edu

We report on a field study demonstrating systematic differences between the preferences people anticipate they will have over a series of options in the future and their subsequent revealed preferences over those options. Using a novel panel data set, we analyze the film rental and return patterns of a sample of online DVD rental customers over a period of four months. We predict and find that *should* DVDs (e.g., documentaries) are held significantly longer than *want* DVDs (e.g., action films) within customer. Similarly, we also predict and find that people are more likely to rent DVDs in one order and return them in the reverse order when *should* DVDs are rented before *want* DVDs. Specifically, a 1.3% increase in the probability of a reversal in preferences (from a baseline rate of 12%) ensues if the first of two sequentially rented movies has more *should* and fewer *want* characteristics than the second film. Finally, we find that as the same customers gain more experience with online DVD rentals, the extent to which they hold *should* films longer than *want* films decreases. Our results suggest that present bias has a meaningful impact on choice in the field, and that people may learn about their present bias with experience and, as a result, gain the capacity to curb its influence.

Key words: *want/should*; intrapersonal conflict; time-inconsistent preferences; present bias; learning

History: Received November 14, 2007; accepted November 7, 2008, by George Wu, decision analysis.

Published online in *Articles in Advance* March 6, 2009.

1. Introduction

Throughout our lives, we face many choices between things we know we *should* do and things we *want* to do: whether or not to visit the gym, whether or not to smoke, whether to order a greasy pizza or a healthy salad for lunch, and whether to watch an action-packed blockbuster or a history documentary on Saturday night. In this paper, we investigate the effects of this type of internal conflict between the desire to do what will provide more short-term utility and the knowledge that it is in our long-term interest to do something else. In particular, we focus on the way this type of conflict leads individuals to make systematically different decisions in the domain of film rentals when they make choices in the present about what to watch versus choices for the future about what to rent.

A number of authors have discussed the distinction between goods that provide primarily long-term benefits, which we call *should* goods, and goods that provide primarily short-term value, which we call *want* goods. Options conceptually similar to *shoulds* have also been called “cognitive,” “utilitarian,” “virtue,” “affect-poor,” and “necessity” options, while options

that are conceptually similar to *wants* have also been called “affective,” “hedonic,” “vice,” “affect-rich,” and “luxury” options (see Khan et al. 2005 for a review). We rely on the terms *should* and *want* to convey the internal tension produced by these competing options. The distinction between these different types of goods is important because evidence suggests that the context in which a decision is made may affect which types of goods, *should* goods or *want* goods, a person prefers.

The tendency to put off options preferred by our *should* selves (e.g., saving, eating vegetables) in favor of options preferred by our *want* selves (e.g., spending, eating ice cream) is stronger for decisions that will take effect immediately than decisions that will take effect in the future (Loewenstein 1996, Bazerman et al. 1998). Economists have modeled this phenomenon by proposing that people dramatically discount future utility relative to present utility (see, for example, Phelps and Pollak 1968, Ainslie 1992, Loewenstein and Prelec 1992, Laibson 1996, O’Donoghue and Rabin 1999), and with “multiple-selves” models in which individuals’ decisions are controlled by multiple internal agents with competing preferences, one of which

optimizes over a longer time horizon than the other and is more likely to control choices that are made for the future than the present (see, for example, Thaler and Shefrin 1981, Read 2001, Fudenberg and Levine 2006). In this paper, we empirically test for the time-inconsistent preferences that these models of present bias predict people will demonstrate when making repeated choices over the same set of goods, ranging from extreme *want* goods (items with only short-term benefits) to extreme *should* goods (items with only long-term benefits), when some decisions will take effect in the present and some will take effect in the future.

Evidence that people prefer *want* options over *should* options more frequently when making choices about the short run rather than the long run has been found in numerous domains (Oster and Scott Morton 2005, Wertenbroch 1998, Rogers and Bazerman 2008, Read and van Leeuwen 1998, Milkman et al. 2008b), including that of film rentals in a laboratory setting (Read et al. 1999, Khan 2007). To extend the study of the impact of present bias on people's preference rankings of *want* and *should* options beyond the laboratory, we obtained a novel panel data set containing individual-level information about consumption decisions over a period of four months from Quickflix®, an Australian online DVD rental company. This data set comes from a domain in which individuals make rental choices for the future and consumption choices for the present from a set of goods that range from extreme *should* items (highbrow films) to extreme *want* items (lowbrow films). Repeated observations of the same individuals over time allow us to investigate both whether customers exhibit present bias and whether they learn to reduce their present bias with experience.

To test the theory that people exhibit present bias in the domain of film rentals, we begin by scoring the films in our data set on the spectrum from *should* to *want* items. We then use our rental data to test and confirm the hypothesis that the same Quickflix customer holds films longer the closer the films fall to the *should* end of the *want/should* spectrum. We also test and confirm the hypothesis that when customers rent two sequential films, the first of which has more *should* and fewer *want* characteristics than the second film, they are more likely to reverse their preferences (watching and returning the films out of order) than when they rent a movie with more *want* and fewer *should* characteristics first. Both of these hypotheses stem from the combination of a model of consumers as present biased and our definition of relative *should* and *want* goods. We thus interpret our findings as evidence that people exhibit present bias in the field when making decisions about film rentals. Finally, we address and attempt to rule out a number

of alternative explanations for our findings. One set of analyses designed to rule out alternative explanations for our results reveals that consumers reduce the extent to which they hold films that fall closer to the *should* end of the *want/should* spectrum longer than other DVDs as they gain rental experience. This suggests that people learn about their present bias over time and that the effects we detect are unlikely the result of "optimal" decision-making strategies.

The remainder of this paper is organized as follows. In §2, we review the relevant literature on time-inconsistent preferences and clarify the origins of our hypotheses. In §3, we describe our data set and methods for rating films along the spectrum from *should* to *want*. We present the results of our analyses and discuss alternative explanations for our findings in §4, and present our conclusions in §5.

2. Past Research on Time-Inconsistent Preferences

A considerable literature on time-inconsistent preferences has developed since Strotz (1956) pointed out that people exhibit more impatience when making decisions that will take effect in the short run rather than the long run. Loewenstein and Thaler (1989), Ainslie (1992), O'Donoghue and Rabin (1999), and Frederick et al. (2002) provide partial reviews of the literature on intertemporal choice, and Milkman et al. (2008a) review the literature on the context effects that have been shown to alter people's preferences for *should* versus *want* options.

Evidence from numerous laboratory studies indicates that consumers exhibit present bias when making choices about money (McClure et al. 2004, Thaler 1981, Kirby and Herrnstein 1995, Kirby and Marakovic 1996, Kirby 1997), lottery tickets (Read et al. 1999), relief from pain and irritation (Solnick et al. 1980, Navarick 1982, Trope and Fishbach 2000), films (Read et al. 1999, Khan 2007), and foods (Wertenbroch 1998, Khan 2007, Read and van Leeuwen 1998), among other things. Models of present bias have also been tested and confirmed in the field in the domains of gym attendance (Malmendier and Della Vigna 2006), magazine newsstand and subscription pricing (Oster and Scott Morton 2005, Wertenbroch 1998), savings behavior (Angeletos et al. 2001, Ashraf et al. 2006), and supermarket quantity discounts (Wertenbroch 1998). Past field studies, however, have not directly tested whether people's preference rankings over a set of goods in advance of consumption are systematically different from those at the time of consumption, as predicted by a combination of a model in which consumers dramatically discount utility from future periods and our definition of *want* and *should* options.

For a number of reasons, it is empirically difficult to test models of consumers as present biased

outside the laboratory. A direct test of any such model requires a data set containing information about the consumption decisions of the same consumers over time, where different decisions take effect at different points in the future. Past field studies have overcome the hurdle of obtaining individual-level consumption data over time in the domains of savings behavior and gym attendance. Partnering with a bank in the Philippines, Ashraf et al. (2006) offered commitment savings products to a subset of the bank's former clients. They confirm (for female subjects) the prediction that consumers who exhibit more present bias on hypothetical questions are more likely to take up commitment devices. Ashraf et al. (2006) track individuals' take up of a savings commitment device as well as the amount those individuals save in their bank accounts over a 12-month period. Malmendier and Della Vigna (2006) employ a panel data set to examine individual-level gym attendance and contract types over a three-year period at several health clubs. They find that present bias explains the popularity of flat-fee contracts among gym customers who could have saved money by paying per visit. Although neither of these studies employs data that would permit the identification of explicit reversals in preferences at the within-subject level, both test predictions of models of present bias in the field at the within-subject level. Both studies also examine the choice of whether or not to engage in a *should* behavior, but not the way in which people dynamically change their preferences over a set of options ranging from those with more *want* characteristics to those with more *should* characteristics, which is the phenomenon examined in this paper.

Tests of the hypothesis that individuals are present biased using between-subjects data are less challenging to perform in the field than tests employing within-subject data. Angeletos et al. (2001) use data from the Panel Study of Income Dynamics to evaluate the relative performance of the competing hyperbolic and exponential time-discount function models. Compared with the exponential discount function, which does not allow for present bias, they find that the hyperbolic discount function, which models consumers as present biased, offers a better approximation of the data on household liquid wealth, credit card borrowing, and changes in consumption in response to predictable changes in income. In another between-subjects field study of present bias, Oster and Scott Morton (2005) examine the newsstand and subscription pricing of *should* and *want* magazines (which they call "meritorious magazines" and "magazines for which consumers might have a time-inconsistency problem," respectively). The authors find, as models of present bias predict, that *should* magazines have a higher subscription-to-newsstand price ratio than

want magazines. Finally, Wertenbroch (1998) examines the quantity discounts applied to a matched sample of 30 virtue (*should*) and 30 vice (*want*) supermarket goods and finds that, consistent with models of present bias in which consumers are assumed to be sophisticated about their self-control problems, *want* goods are, on average, subject to steeper quantity discounts than *should* goods. He also estimates the price elasticity of demand for a sample of paired vice (*want*) and virtue (*should*) groceries using a year of supermarket scanner data. Again, consistent with models of present bias, Wertenbroch (1998) finds that demand for *should* goods is more price sensitive than demand for *want* goods. However, all of these studies are tests of the implications of models of present bias on outcomes that are one or more levels removed from individuals' actual choices.

In our study, we attempt to combine the approaches of Wertenbroch (1998) and Oster and Scott Morton (2005), who examine the implications of models of present bias in field domains where consumers are faced with ranking their preferences over a range of goods, with the approaches of Ashraf et al. (2006) and Malmendier and Della Vigna (2006), who use within-subject data sets to test various predictions of models of present bias in the field. The central hypotheses of this paper and the domain of interest were inspired by Read et al. (1999), who conduct a laboratory experiment to show that when choosing a film for immediate consumption, people more often prefer movies with more *want* characteristics and fewer *should* characteristics than when selecting a film for delayed consumption. Others have hypothesized that online DVD rental customers might exhibit a tendency to hold highbrow films longer than lowbrow films (see Phillips 2006, Tugend 2006, Goldstein and Goldstein 2006), but none have presented empirical support for their conjectures. Our goal is to provide the first direct, within-subject field test of whether consumers exhibit present bias in a domain where their choice set includes options ranging from *want* to *should* items. In addition, we look for evidence that customers learn about their present bias as they gain experience.

Models of present bias suggest that, given a set of options, an option that provides more long-term benefits (a *should* option) will be relatively more attractive than an option that provides more short-term benefits (a *want* option) when a choice is made for the future than when that same choice is made for the present. Because the decision of which film from a collection to watch first (and thus return first) is a choice made over a set of options for the present, theories of present bias lead to the following hypothesis:

HYPOTHESIS 1A. *The closer a film falls to the should end of the want/should spectrum, the longer a customer will postpone watching and thus returning it.*

Similarly, because the decision of which film to rent is a choice made for the future but the decision of which film to watch (and thus return first) is made for the present, theories of present bias lead to the following hypothesis:

HYPOTHESIS 1B. *The probability that a customer will return two sequentially rented films out of order increases as the first film rented becomes more of a should on the want/should spectrum relative to the second film.*

Finally, in an attempt to rule out the possibility that individuals have a rational reason for holding more extreme *should* films longer than other films, we look for evidence that customers learn to reduce the extent to which they exhibit this tendency as they gain experience with online DVD rentals. If it were optimal for customers to hold more extreme *should* films longer than others, we would not expect experience to diminish this effect. Past research on commitment devices has shown that some people who exhibit present bias are sophisticated about their self-control problems and willing to incur costs to reduce the effects of their present bias (Wertenbroch 1998, Ariely and Wertenbroch 2002, Ashraf et al. 2006). These results raise the question of whether people gain sophistication about their present bias through experience or whether sophistication is a stable trait. It has been demonstrated in a number of domains that people have the ability to learn from experience to reduce their decision-making errors (Lichtenstein and Fischhoff 1980, Erev and Roth 1998). As we address potential alternative “rational” explanations for our results, we look for evidence that as customers gain experience renting DVDs, they reduce the extent to which they procrastinate about watching films that fall closer to the *should* end of the *want/should* spectrum.

3. Methods

3.1. Data Set

We obtained a novel panel data set from Quickflix, the second largest online DVD rental company in Australia, containing information about the individual choices made by the company’s customers between March 1, 2006, and June 30, 2006. Customers in the most popular Quickflix subscription plan pay a flat fee each month to hold three exchangeable films in their homes, and they may hold the movies they rent for an unlimited length of time without incurring late fees. To ensure that all customers in our data set are subject to the same incentives with regard to their film rental and return behavior, we conduct our analyses only on customers in Quickflix’s most popular plan. Quickflix offers a selection of over 15,000 movie titles, and each customer maintains a “queue,” or an ordered list of the movies she would like to rent. When a customer

returns one film, Quickflix immediately sends that customer the film listed at the top of her queue. When a customer’s first choice is unavailable, the next highest film in her queue is sent instead. For a typical subscriber, the net turn around time for a film exchange is two days, and postage is paid by Quickflix.

Our data set includes the day-to-day records of people’s film rentals and returns over a four-month period. Although our rental data set ends on June 30, 2006, we have records of the dates when each of the films rented during the relevant time period was returned. Quickflix also provided us with unique identifiers for each customer and with descriptive information about each film in its database. During the four-month period included in our data set, a total of 4,474 different customers participated in Quickflix’s most popular three-at-a-time unlimited DVD rental plan, renting a total of 101,545 DVDs (an average of 22.7 per customer). On average, these customers held the DVDs they rented for 12 days, and 90% of movies were held between 4 and 32 days.

3.2. Assigning Films Continuous *Should/Want* Scores

To test our hypotheses about how films’ positions on the *should/want* spectrum affect both (a) the order in which they are returned relative to the order in which they were rented and (b) how long it takes customers to return them, we must create a measure of the extent to which a film is a *should* versus a *want*. We first create separate scores for films on a *should* spectrum and a *want* spectrum and then subtract films’ *want* scores from their *should* scores to measure where each film fits on the spectrum from an extreme *should* option to an extreme *want* option.¹

To generate a measure of each Quickflix film’s *should minus want* (SMW) score, we borrow data from a previous research project. For that project, 145 anonymous American volunteers who signed up to participate in online paid polls administered by Harvard Business School’s Computer Lab for Experimental Research were paid \$15 to give *should* and *want* scores to a random sample of 60 films from a database of 1,040 movies. Raters ranged in age from 18 to 45, with an average age of 25, and 70% of raters were female. After being provided with concept definitions, subjects in this study were first asked to give 60 films *want/(should)* ratings ranging from 1 to 7 and

¹ Through pretesting, we determined that the extent to which a film is a *should* film (based on how much long-term value it provides) is most easily evaluated distinctly from the extent to which it is a *want* film (based on how much short-term value it provides). However, the variable of interest in this study is where on the spectrum from an extreme *should* to an extreme *want* a film lies, which theories of present bias suggest will predict the extent to which it is preferred when choices are made for the future versus the present.

were then asked to give the same set of 60 films *should/(want)* ratings ranging from 1 to 7 (see the online appendix, provided in the e-companion,² for more details). The order in which subjects were asked to rate the 60 films was randomized, as was the sequence in which they provided *should* and *want* ratings (50% gave films *should* ratings first). Subjects saw the same information about a film that they would have seen if they had searched for it on the website of the American online DVD rental company Netflix. We provided subjects with an incentive to provide accurate ratings of films by paying them for performance. For each film a survey participant classified within one point of the average rating across respondents, her “accuracy score” was increased by one. The 20% of participants who received the highest accuracy scores received a \$10 bonus payment.

Five hundred of the 1,040 movies from this survey were also films in the Quickflix DVD rental database. Because the films subjects rated were randomly selected, these 500 films were rated by varying numbers of subjects. An average of 8.58 survey participants rated each film (standard deviation = 3.02).

To confirm that subjects provided us with reliable ratings of the movies in our survey database, we conduct an analysis of inter-rater reliability. Because we are interested in quantifying each film’s *should minus want score*, we first calculate this difference variable for each film-rater pair. If our survey ratings contain a meaningful signal, the *SMW scores* assigned by different survey participants to the same film should be more tightly clustered than the *SMW scores* assigned by different survey participants to different films. We run a one-way analysis of variance to compare ratings variation between films with ratings variation within films (Shrout and Fleiss 1979). An intraclass correlation of 0.21 and an estimated reliability of a film *should minus want score* mean of 0.70 confirms that our survey averages are reliable: *should minus want scores* vary significantly more between films than within films. To check that participants understood from our definitions of *want* and *should* movies that, for the most part, extreme *should* movies are not extreme *want* movies and vice versa, we examine the correlation between a movie’s average *want* score and its average *should* score. This correlation is highly significant and negative ($\rho = -0.22$, $p < 0.001$) across the 500 films in our sample, suggesting that our raters grasped the relationship between a typical movie’s *want* and *should* characteristics.

To validate these scores and ensure that the incentives we provided to subjects for performance did not

bias their ratings, we hired five research assistants to assign each of the 500 films in our sample *want* and *should* scores. These research assistants were provided with the same concept definitions as our original subjects and the same information about each film, and the order in which they rated films was randomized. Our research assistants were each paid a flat fee of \$120 for their time with no bonus pay for “accuracy.” The Cronbach’s alpha across these five raters’ *should minus want scores* for the 500 films in our sample was 0.64, indicating a good level of agreement (Nunnally 1967). The correlation between the average *should minus want scores* produced by our original 145 subjects and our five research assistants across the 500 films in our database was 0.68 ($p < 0.001$). This high correlation between the two sets of ratings gives us confidence that paying subjects for performance did not harm the reliability of the survey data we collected on films’ *should minus want scores*. It is also worth noting that none of the primary regression results presented in this paper differ meaningfully in magnitude or significance if we rerun our analyses using the ratings data obtained from research assistants instead of the data obtained from our online survey respondents.

To develop a *should minus want score* for each film in the Quickflix DVD library using the survey data from 145 subjects described above, we ran an ordinary least squares (OLS) regression to predict survey respondents’ average *should minus want scores* of the 500 Quickflix films from our survey (see Table 1). We employed analytic weights in our regression to control for the fact that different numbers of subjects rated each film. The predictors in our regression include all of the quantifiable characteristics of a film that were provided to us by Quickflix: 21 genre dummy variables, the average subscriber’s rating of the film, the number of years since the film was released in theatres, the number of days since the film was released on DVD, dummy variables indicating the film’s OFLC rating,³ the number of characters in the film’s title, and the number of other films in the Quickflix rental database that were released by the same studio. Regression (1) in Table 1 explains 44% of the variance in films’ average *should minus want scores*.

We extrapolate from our sample of 500 films to the 17,258 films in the Quickflix movie database and give each film an *SMW score* based on the coefficients in regression (1). According to this classification scheme, the movie with the lowest *SMW score* in our database is *The Story of Ricky*, a violent, futuristic, sci-fi horror film from 1991, and a film that seems intuitively likely

² An electronic companion to this paper is available as part of the online version that can be found at <http://mansci.journal.informs.org/>.

³ The Australian Office of Film and Literature Classification (OFLC) provides films with ratings based on the offensiveness of their content (<http://www.oflc.gov.au>).

Table 1 Predicting a Movie's Average *SMW Score*

	Average <i>SMW</i> score from survey (1)
Intercept	-0.6128
Number of characters in film title $\times 10^2$	-0.5069
Number of films released by same studio in Quickflix library $\times 10^3$	0.6732
Average user rating of film $\times 10^2$	0.5435
Years since film's release in theaters	0.0216***
Days since film's release on DVD $\times 10^4$	0.5400
Film rated R	-1.4142***
Film rated PG	-1.1275**
Film rated MA	-0.9434**
Film rated M15	-0.9854
Film rated M	-1.1670***
Film rated G	-0.9747**
Action film	-0.3636**
Adventure film	-0.0448
Anime film	-1.0729***
Arthouse film	0.0331
Australian film	1.1523
Children's film	-0.7641**
Comedy film	-0.8284***
Crime film	-0.7289***
Documentary film	1.9466***
Drama film	0.4830***
Family film	-0.5135**
Fantasy film	-0.2983
Foreign film	0.3406
Horror film	-1.0109***
Lifestyle film	0.1081
Music film	-0.0772
Performance film	-0.8155
Romance film	-0.5658***
Science fiction film	-0.6727***
Sports film	-0.5500
Television film	-0.4422**
Thriller film	-0.5886***
Analytic weights	Yes
Observations	500
R^2	0.4392

Notes. Column (1) reports the OLS coefficients from regressions of the average *should minus want scores* of films on various attributes of each film.

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

to be a strong *want* for anyone who would choose to rent it. The movie with the highest *SMW score* in our database is *Kokoda Frontline*, an Australian, Oscar-winning documentary from 1942 about the Kokoda campaign in Papua New Guinea during World War II, a film that seems intuitively likely to be a strong *should* for anyone who would choose to rent it. We then standardize these scores across the 17,258 films in the Quickflix library.⁴

⁴ If we replicate these procedures to develop a standardized *should minus want score* using the second set of film ratings discussed above, which were provided by five research assistants, the correlation between the two sets of ratings across the 17,258 films in our database is 0.9341 ($p < 0.0001$).

Table 2 The Effect of a Movie's *SMW Score* on Holding Time

	Log(holding time) (2)
<i>Should minus want score</i>	0.0206*** (0.0018)
Movie's rank in customer's queue when shipped	-0.0002* (0.0001)
Days movie spent in customer's queue	0.0002*** (0.0000)
Customer's DVD rentals since January 1, 2006	0.0017*** (0.0002)
Movie's length (in minutes)	0.0001*** (0.0000)
Day of the week movie shipped fixed effects	Yes
Week of the year fixed effects	Yes
Customer fixed effects	Yes
Observations	101,545
Customers	4,474
R^2	0.5294

Notes. Column (2) reports OLS coefficients from a regression of the log of the number of days a customer held a movie on the movie's *should minus want score*, controlling for the other variables listed. Robust standard errors clustered at the customer level are in parentheses.

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

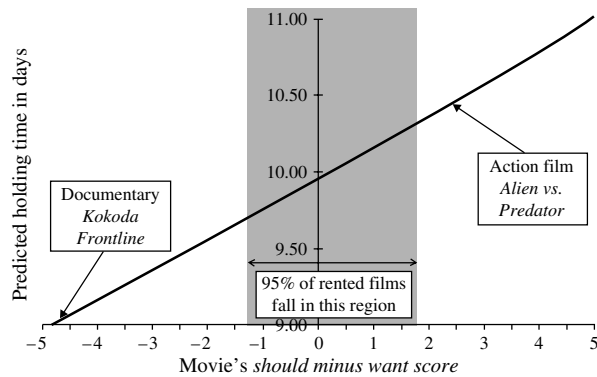
4. Results

4.1. Holding Time

We address the question of whether or not Quickflix customers exhibit the behavior predicted by models of present bias by running a series of regressions. First, we examine the influence of a movie's *should minus want score* on how many days a customer holds that film. A combination of a model of present bias with our definitions of relative *want* and *should* goods suggests that the higher a movie's *should minus want score*, the more likely a customer will be to postpone watching it, leading her to hold it longer. In Table 2, we present the results of an OLS regression estimating the relationship between the logarithm of how many days a customer holds a movie before returning it and that movie's *should minus want score*.⁵ In this regression, the explanatory variables include a measure of the movie's *should minus want score*, the rank of the movie in a customer's queue when it was shipped, the number of days the movie spent in the customer's queue before it was rented, the number of movies the customer had rented from Quickflix since January 1, 2006, when the movie was shipped, the length of the movie in question, dummies indicating

⁵ Our outcome variable is the logarithm of a movie's holding time rather than the raw holding time because it seems more appropriate to assume that consumers increase the relative rather than the absolute holding time of a film based on its position along the *should/want* spectrum. However, our findings are robust to examining raw holding time.

Figure 1 Illustration of the Relationship Between a Film's *SMW Score* and Its Predicted Holding Time



the day of the week when the movie was shipped, and dummies indicating the week of the year when the movie was shipped. This regression also includes customer fixed effects, and standard errors are clustered by customer.

The coefficient on the *SMW score* of a film in regression (2) (Table 2) indicates that, holding all else constant, a one standard deviation increase in a movie's *SMW score* is associated, on average, with a 2% within-customer increase in how many days the movie is held. To put this in context, the results from regression (2) indicate that for the same customer, *Kokoda Frontline* will be held, on average, 17% longer (an average of 1.5 days longer) than *Alien vs. Predator*, a lowbrow 2005 action, sci-fi thriller that received one of the lowest *should minus want scores* of the 17,258 films in the Quickflix library. Figure 1 illustrates this result graphically.

4.2. Reversals in Preferences

Next, we test the prediction that people are more likely to rent one movie before another but reverse the order in which those movies are watched (and thus returned) when the first movie rented receives a higher *should minus want score* than the second. To test this prediction, we create a data set in which each observation corresponds to an instance in which a Quickflix customer rented two movies, one after the other, and the second movie was delivered to that customer at least one day before the first had been received by a Quickflix return center. We only include such an event in our data set if the first movie had a higher ranking in the customer's queue when it was mailed than the second movie, an indication that the customer explicitly preferred the first movie to the second when making a decision that would take effect in a future period about which movie to watch. We also exclude all observations where the first rental in a pair of sequential rentals was the second rental in a previous pair (to avoid including correlated observations). Of the 11,964 sequential rentals in our resulting

Table 3 The Effect of a Movie's *SMW Score* on Reversals in Preferences

	Preference reversal (3)	Preference reversal (4)
Movie 1's <i>should minus want premium</i>	0.008*** (0.003)	
Movie 1 received a higher <i>SMW score</i> than movie 2		0.013** (0.007)
Movie 1's length in minutes minus movie 2's length in minutes $\times 10^3$	0.018 (0.028)	0.018 (0.028)
Customer fixed effects	Yes	Yes
Observations	11,964	11,964
Customers	3,079	3,079
R^2	0.327	0.326

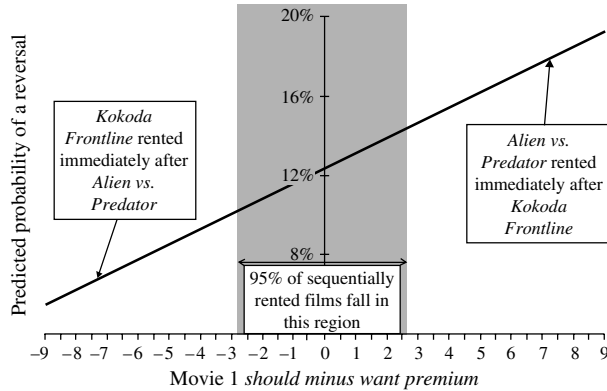
Notes. Columns (3) and (4) report OLS coefficients from regressions to predict whether or not a customer exhibited a reversal in preferences based on different measures of the *should minus want scores* of movies in the customer's choice set. Robust standard errors clustered at the customer level are in parentheses.

*, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

data set, we observe 1,478 reversals in preferences. To create a measure of how much more of a *should* movie and less of a *want* movie the first film rented is relative to the second film, we subtract the *SMW score* of the movie that was rented second (movie 2) from the *SMW score* of the movie that was rented first (movie 1). We call this variable the *movie 1 SMW premium*. Its mean value in our sample of sequentially rented movies is -0.001 , and its standard deviation is 1.20. Of the 6,019 sequential rentals in which the *movie 1 SMW premium* is positive, we observe reversals in preferences at a rate of 13.3%, and of the 5,945 sequential rentals in which the *movie 1 SMW premium* is negative, we observe reversals in preferences at a rate of 11.3%.

In Table 3, we present the results of two of OLS regressions estimating the relationship between the probability of an intertemporal reversal in preferences and the *movie 1 SMW premium*, controlling for the difference between the films' lengths and customer fixed effects, and clustering standard errors by customer. Regression (3) demonstrates that for each additional standard deviation by which movie 1's *SMW score* exceeds that of movie 2, the probability of a reversal in preferences increases by 0.8% (an approximately 7% increase from the mean 12% probability of a reversal in preferences). See Figure 2 for an illustration of this effect. In regression (4), we replace our continuous measure of movie 1's *should minus want premium* over movie 2 with a dummy variable indicating whether or not movie 1's *SMW score* exceeds that of movie 2. Regression (4) demonstrates that when movie 1's *SMW score* exceeds movie 2's, the probability of a reversal in preferences is 1.3% higher than it

Figure 2 Illustration of the Relationship Between the Movie 1 SMW Premium When Movie 1 Is Rented Immediately Before Movie 2 and the Predicted Probability That a Customer Will Reverse Her Preferences and Return Movie 2 Before Movie 1



would be otherwise (an approximately 11% increase from the mean 12% probability of a reversal in preferences). Running these analyses with a conditional logit model including customer fixed effects and clustered standard errors yields nearly identical results.

The increases in the likelihood of a reversal in preferences reported in Table 3 are conservative estimates of the effect we seek to quantify because many of the data points included in our analyses may not coincide with situations in which a customer had an actual opportunity to reverse her preferences. In some cases, movie 2 must have arrived in a customer’s home after movie 1 had been watched or even mailed back to Quickflix.⁶ Each point included in our data set that corresponds to a situation in which a customer did not have an opportunity to reverse her preferences necessarily reduces the change in the probability of a reversal in preferences we are able to associate with a change in the *movie 1 SMW premium*.

4.3. Comparing Our *Should Minus Want Score* to Other Available Predictors of Holding Time

We next investigate the extent to which our *should minus want score* captures the key components of a film that predict how long a Quickflix customer will

⁶ Because we know only the dates when return centers received films from customers and the dates when they shipped new movies to customers, and not how long shipments spent in the mail, the best we can do is to assume all shipments spent one day in the mail, which is almost certainly an underestimate. Our finding that the probability of a preference reversal increases as the *movie 1 SMW premium* increases is sensitive to this assumption: the coefficient on our main effect increases, for example, the more days we assume shipments spend in the mail. However, such assumptions about shipment time, when false, might bias our results upward by selecting on films that have a tendency to spend longer in customers’ homes, so we make the most conservative possible assumption about shipping time in the analyses we present.

Table 4 The Effect of a Movie’s Quantifiable Attributes on Holding Time

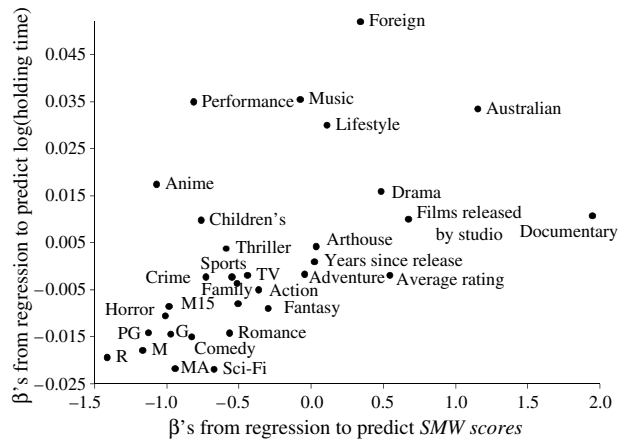
	Log(holding time) (5)
Number of characters in film title × 10 ²	-0.0079
Number of films released by same studio in Quickflix library × 10 ³	0.0129
Average user rating of film × 10 ²	-0.0019
Years since film’s release in theaters	0.0010***
Days since film’s release on DVD × 10 ⁴	0.0787***
Film rated R	-0.0194
Film rated PG	-0.0142
Film rated MA	-0.0217
Film rated M15	-0.0086
Film rated M	-0.0179
Film rated G	-0.0144
Action film	-0.0050
Adventure film	-0.0017
Anime film	0.0174
Arthouse film	0.0042
Australian film	0.0334***
Children’s film	0.0098
Comedy film	-0.01496***
Crime film	-0.0023
Documentary film	0.01069
Drama film	0.0159***
Family film	-0.0037
Fantasy film	-0.0090
Foreign film	0.0520***
Horror film	-0.0105
Lifestyle film	0.0299*
Music film	0.0354***
Performance film	0.0349*
Romance film	-0.0142**
Science fiction film	-0.0219***
Sports film	-0.0023
Television film	-0.0020
Thriller film	0.0038
Movie’s rank in customer’s queue when shipped	-0.0003**
Days movie spent in customer’s queue	0.0001***
Customer’s DVD rentals since January 1, 2006	0.0017***
Movie’s length (in minutes)	0.0001***
Day of the week movie shipped fixed effects	Yes
Week of the year fixed effects	Yes
Customer fixed effects	Yes
Observations	101,545
Customers	4,474
R ²	0.5301

Notes. Column (5) reports the OLS coefficients from regression of the log of the number of days a customer held a movie on various attributes of that movie. Robust standard errors clustered at the customer level.

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

postpone watching it. Regression (5) in Table 4 presents the result of an OLS regression conducted to predict the logarithm of the number of days a customer will hold a movie before returning it using the same predictors that were used in regression (1) to predict the *should minus want scores* that survey participants gave to a sample of 500 Quickflix movies. The regression in Table 4 includes the same

Figure 3 Illustration of the Correlations Between the β Coefficient Estimates in Regressions (1) and (5)



control variables, fixed effects, and clustered standard errors that were used in regression (2). The correlation between the coefficients on each of the movie descriptor variables used as predictors in regression (5) and the coefficients on each of these same predictor variables in regression (1) is 0.53 (see Figure 3 for illustration). This suggests that our *SMW scores* capture the essence of those characteristics of a DVD that predict how long a Quickflix customer will postpone watching it.

4.4. Customer-Level Analyses

In addition to examining average levels of present bias across customers by including all individuals in our database in regressions with customer fixed effects and clustered standard errors, we replicate our primary analyses at the customer level. This allows us to determine what percentage of customers appear to exhibit present bias as well providing a second measure of the magnitude of the effects of interest.

In our first set of customer-level analyses, for each of the 4,474 customers included in the holding time regressions described in §4.1, we run a regression to predict the logarithm of the number of days a customer holds a movie before returning it with a single predictor variable: the movie's *SMW score*. Of the 3,915 customers who rented enough movies for us to estimate a beta coefficient and associated standard error for the single predictor variable in their OLS regression, the coefficient estimated on the *SMW score* variable is positive 55% of the time (binomial test, $N = 3,915$, $H_0: 55.3\% = 44.7\%$ can be rejected, $p < 0.001$). The weighted average beta estimate on our primary predictor variable (weighted by the inverse of a coefficient estimate's standard error to account for differences in the precision of each beta estimate) is 0.047. The average number of observations included in these 3,915 regression equations was 25.7 (standard deviation = 20.9).

In our next and final set of customer-level analyses, for each of the 3,079 customers included in the regressions examining reversals in preferences described in §4.2, we run a regression to predict the probability of a reversal in preferences at the individual level with a single predictor variable: the *movie 1 SMW premium*. Of the 818 customers who exhibited enough reversals in preferences for us to estimate a beta coefficient and associated standard error for the single predictor variable in their OLS regression, the coefficient estimated on the *movie 1 SMW premium* is positive 55% of the time (binomial test, $N = 818$, $H_0: 54.8\% = 45.2\%$ can be rejected, $p < 0.01$). The weighted average beta estimate on our primary predictor variable (again, weighted by the inverse of a coefficient estimate's standard error to account for differences in the precision of each beta estimate) is 0.058. The average number of observations included in these 818 regression equations was 6.4 (standard deviation = 3.0).

The results of these analyses are consistent with the findings we presented in §§4.1 and 4.2 and provide alternative effect size estimates. The effect sizes estimated here are two and a half to seven times as large as those estimated in our fixed effects regressions. These analyses also give us a sense of the percentage of customers in our data who exhibit present bias, although it is important to note that the customers included in these analyses are, on average, more frequent renters than those included in our primary analyses, which may bias these numbers.

4.5. Addressing Alternative Explanations for Our Findings and Seeking Evidence of Learning

In addition to present bias, there are a number of potential alternative explanations for our findings, which we will attempt to rule out in this section. The first is that movies with higher *SMW scores* are also movies that people like more. As a result of this, people hold onto these types of movies for longer periods of time to watch them repeatedly, lend them to friends, or draw out the viewing experience. There are several reasons we believe we can rule out this explanation for our results. First, assuming a movie's popularity (which we quantify as the number of times it was rented in our data set divided by the number of days when it was available for rent) is a reasonable proxy for how well people like a movie, if this alternative explanation were correct, we would expect a movie's *SMW score* to be positively correlated with its popularity. In fact, we find that this popularity measure is significantly negatively correlated with a movie's *SMW score* across the 17,258 movies in our data set ($\rho = -0.11$, $p < 0.001$).

Other ways to examine the plausibility of this alternative hypothesis are to see if our primary results change if we (a) include popularity as a predictor in

our primary regression analyses or (b) restrict our primary regression analyses so they only include observations involving popular movie rentals. When we add the popularity measure described above to the set of explanatory variables included in regression (2) to predict a movie's holding time, we find that a film's popularity is significantly negatively associated with its holding time, and the addition of this variable does not meaningfully affect the estimated coefficient on a movie's *SMW score* or its statistical significance. Similarly, when we run regression (2) to predict a movie's holding time and restrict our sample to include rentals of only the 200 most popular Quickflix movies, our main effect remains significant, and our estimate of its effect size increases by over 30%. Turning to our analyses of reversals in preferences, including a variable that quantifies the popularity premium of the first of two sequentially rented movies when we run regression (3) to predict the probability of a reversal in preferences also has no meaningful effect on our estimate of the coefficient on our primary predictor variable (the *movie 1 SMW premium*) or on its statistical significance. In addition, movie 1's popularity premium is negatively related to the probability of a reversal in preferences. Finally, restricting our analysis of reversals in preferences to include only particularly popular film titles again increases the estimated beta on the *movie 1 SMW premium*, this time by nearly 80%.⁷ Together, these results suggest that our findings are not driven by people's preference for *should* movies and resulting tendency to hold onto these types of films for longer.

Another potential alternative explanation for our results is that people rarely find themselves in the mood for a *should* movie, but when they do, having a *should* movie on hand is extremely valuable. As a result, people hold *should* movies longer than *want* movies, but it is rational for them to do so because of the high "option value" of these films. If this is truly the source of our finding that *should* films are held longer than *want* films, and if it is not a rationalization people provide for their tendency to procrastinate about watching *should* films, then we would not expect to see the same customer with more experience renting from Quickflix attenuate her tendency to hold *should* films longer than *want* films. However, if what we are observing is irrational procrastination, customers with more DVD rental experience ought to learn about their present bias and take steps to curb it.

To pit the alternative explanations for our primary findings outlined above against one another, we

⁷ In this case, we restrict ourselves to the top 1,500 films because we lose too many observations to estimate within-person effects with any precision if we restrict ourselves to observations involving two top-200 films rented sequentially.

Table 5 The Effect of a Customer's Experience on How Much a Movie's *SMW Score* Affects Its Holding Time

	Log(holding time) (6)
<i>Should minus want score</i>	0.0208*** (0.0018)
(<i>SMW score</i>) × (Z Customer's DVD rentals since January 1, 2006)	−0.0079*** (0.0017)
Z Customer's DVD rentals since January 1, 2006	0.02951*** (0.0042)
Movie's rank in customer's queue when shipped	−0.0002 (0.0001)
Days movie spent in customer's queue	0.0002*** (0.0000)
Movie's length (in minutes)	0.0001*** (0.0000)
Day of the week movie shipped fixed effects	Yes
Week of the year fixed effects	Yes
Customer fixed effects	Yes
Observations	101,545
Customers	4,474
R ²	0.5295

Notes. Column (6) reports OLS coefficients from a regression of the log of the number of days a customer held a movie on the movie's *should minus want score* and the interaction of this variable with the number of films the customer has rented since January 1, 2006, controlling for the other variables listed. Robust standard errors clustered at the customer level are in parentheses.

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

run a new regression to predict a movie's holding time in which we interact the (standardized) number of DVDs a customer has rented from Quickflix since January 1st, 2006 (our proxy for "experience" with DVD rentals), with a movie's *SMW score* (see Table 5, regression (6)). The significant negative coefficient on the interaction term in regression (6) indicates that the more experience a customer has renting DVDs from Quickflix, the less that customer will procrastinate about returning *should* films.⁸ These results are robust to including only observations involving rentals of the 200 most popular Quickflix movies. In addition, customer-level regressions run on customers who rented more than 20 DVDs during the period examined (giving them adequate opportunities for learning) and that include the predictors *SMW score*, rentals year to date, and the interaction

⁸ Running the same type of analysis with our preference reversal regression specification yields a coefficient on the interaction between the number of rentals year to date when movie 1 is rented and the *movie 1 SMW premium* that is directionally consistent with this story (more rentals attenuate the impact of the *movie 1 SMW premium* on the probability of a preference reversal), but this effect is not statistically significant. However, these regressions have considerably less power than our holding time analyses because of the zero-one outcome variable and reduced sample size.

between these two terms also yield a negative interaction coefficient significantly more often than not (binomial test, $N = 1,900$, H_0 : 51.4% = 48.6% can be rejected, $p < 0.05$). It is also worth noting that the newest Quickflix customers exhibit a steeper learning curve than older Quickflix customers,⁹ and that customers tend to exhibit more variance in film holding time the longer they have been Quickflix subscribers (so our findings in Table 5 cannot be explained by a simple reduction in holding time variance with experience story). Together, the above results suggest that customers learn about their present bias and take successful steps to curb their tendency to hold *should* films longer than *want* films as they gain experience renting DVDs. If customers were holding *should* DVDs longer than *want* DVDs for a rational reason (such as the high option value of *should* movies), we would not expect to see this pattern of “learning.” We thus believe we can rule out this alternative “rational” explanation for our results with the aforementioned evidence, which suggests that it may be possible for individuals to learn improved self-control. It is also interesting to note that customers actually rent films with slightly higher *SMW scores* as they gain more rental experience, although this trend is insignificant. This suggests that customers are not learning to stop renting the types of films that they procrastinate about watching, but that instead customers are learning strategies to prevent themselves from exhibiting so much present bias.

To further assess the plausibility of the alternative explanations for our primary findings discussed above, we also conducted a survey in which we asked people with experience renting DVDs what they think has caused them in the past to exhibit the type of behavior we detected in our primary analyses. One hundred and twenty-one subjects who signed up to participate in online paid polls conducted by the market research firm Zoomerang gave a response other than “not applicable” to a question asking why, if they had ever rented a *should* movie before a *want* movie but returned the movies out of order, they thought they did so. Of the three possible explanations subjects could select, just 4.1% of respondents believed they had exhibited this behavior because they “liked the *should* DVD so much more than the *want* DVD that they held onto it longer.” However, 64.5% of respondents claimed they “watched the *want* DVD first because when the moment to choose a DVD to watch arose, the *want* DVD was just more appealing than

the *should* DVD”—an explanation suggesting that present bias drives the time-inconsistent preferences we observe in our data set. The remaining 31.4% of respondents believed they “watched the *want* DVD first because [they were] holding the *should* DVD on hand so that [they] would have it available for when [they were] in the mood to watch it,” an explanation consistent with a rational “option value” story or with the possibility that present-biased renters are poor forecasters of their future moods but naïve about this weakness. These survey results provide additional evidence that the time-inconsistent preferences we observe in our Quickflix data set result primarily from present bias. When asked to explain our findings, the vast majority of subjects with online DVD rental experience point to the explanation we classify as “procrastination” (binomial test, $N = 121$, H_0 : 64.5% = 35.5% can be rejected, $p < 0.001$).

One could construct other explanations for our results in addition to those addressed above and our explanation that renters are present biased. However, in light of the considerable body of research suggesting that people are present biased, including the laboratory study by Read et al. (1999) of present bias and movie choice, and in light of the analyses presented above to rule out alternative explanations, we believe that present bias is the most compelling explanation for our findings.

5. Discussion

The results presented above demonstrate that, consistent with the combined predictions of models of present bias and our definition of relative *should* and *want* goods, the more *should* characteristics and the fewer *want* characteristics a DVD has, the longer a Quickflix customer will postpone watching that DVD. Also consistent with models of individuals as present biased and our definitions of relative *want* and *should* goods, the probability that a customer will exhibit a reversal in preferences increases, when two films are rented sequentially the relatively more *should* and fewer *want* characteristics the first film rented has than the second.

Our analyses offer the first field demonstration that combining a model of consumers as present biased with a model of goods ranging from extreme *wants* to extreme *shoulds* correctly predicts the way people’s rankings of a series of goods will change when they choose for the present versus the future. Our findings are consistent with past work on time-inconsistent preferences, but also extend previous field studies of present bias and choice, which have only examined whether people are less likely to engage in a *should* option when choosing for now versus later. Our evidence suggests that the effects of present bias on which

⁹ We know the order in which Quickflix customers joined the service, and we observe that the 25% of customers who most recently adopted the service exhibit a steeper learning curve than the 50% of customers who most recently adopted the service, who in turn exhibit a steeper learning curve than the entire customer population.

alternative people will select given an array of choices can be quite meaningful in the field. We believe this discovery should increase the importance researchers assign to the results of previous laboratory studies about the effects of present bias on decision making.

In addition, our analyses provide early evidence that experience may attenuate people's degree of present bias. One implication of this finding is that parties with more experience making decisions in a certain domain may better recognize people's tendency to exhibit present bias when making choices in that domain and may take advantage of this "weakness" in their less-experienced counterparts. For example, baseball team managers with extensive experience negotiating players' contracts may realize that players are present biased and value up-front pay (i.e., signing bonuses) more than their yearly salary, an observation that could allow those managers to structure deals that are psychologically more attractive to players, yet less beneficial to them from a normative perspective, and more attractive to the sophisticated managers. Policy makers might want to prevent experienced participants in markets who are sophisticated about present bias from taking advantage of less-experienced, present-biased individuals.

Our findings also have implications for companies that loan items to consumers. Such companies should be able to forecast how long customers will hold different items they have borrowed based on the extent to which those items are *should* versus *want* goods. Specifically, we believe rental companies would be wise to expect customers to return *want* goods faster than *should* goods. Our results have similar implications for online and catalogue retailers that offer different shipment options to customers. Our findings suggest that *want* goods will be in higher demand for immediate delivery than for delayed delivery, whereas *should* goods will exhibit the opposite demand pattern. Finally, our study has implications for consumers, many of whom may be doing an ineffective job of maximizing their utility over time due to their impulsivity. Such consumers would presumably benefit from becoming aware of their present bias, because this would allow them to take steps to curb harmful impulsive behaviors.

6. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://mansci.journal.informs.org/>.

Acknowledgments

The authors gratefully acknowledge helpful and insightful input from two anonymous referees; N. Ashraf, J. Beshears, D. Chugh, S. Hanson, D. Laibson, G. Loewenstein, K. McGinn, D. Moore, S. Mullainathan, D. Parkes,

W. Simpson, A. Sunderam, S. Woolverton; members of Max's Non-Lab; and seminar participants at Harvard, Yale, Stanford, Cornell, the University of Chicago, London Business School, Northwestern, Columbia, the University of California at Berkeley, and the 2008 Society for Judgment and Decision Making Conference. In addition, the authors are thankful to Z. Sharek, who helped them design and implement their survey to measure a film's *should minus want score*; to L. Anik and M. Norton, who helped them run their survey to ask people why they hold *should* films longer than *want* films; and to K. Chance, M. McCoy, J. Packi, E. Siegle, and E. Weinstein, who helped them by giving films *want* and *should* ratings. They also appreciate the cooperation of their contacts at Quickflix (<http://www.quickflix.com.au>) and the help of A. Elberse in facilitating their relationship with Quickflix.

References

- Ainslie, G. 1992. *Piconomics*. Cambridge University Press, Cambridge, UK.
- Angeletos, G., D. Laibson, A. Repetto, J. Tobacman, S. Weinberg. 2001. The hyperbolic consumption model: Calibration, simulation, and empirical evaluation. *J. Econom. Perspectives* 15(3) 47–68.
- Ariely, D., K. Wertenbroch. 2002. Procrastination, deadlines, and performance: Self-control by pre-commitment. *Psych. Sci.* 13(3) 219–224.
- Ashraf, N., D. Karlan, W. Yin. 2006. Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *Quart. J. Econom.* 121(2) 635–672.
- Bazerman, M. H., A. E. Tenbrunsel, K. A. Wade-Benzoni. 1998. Negotiating with yourself and losing: Making decisions with competing internal preferences. *Acad. Management Rev.* 23(2) 225–241.
- Erev, I., A. Roth. 1998. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *Amer. Econom. Rev.* 88(4) 848–881.
- Frederick, S., G. Loewenstein, T. O'Donoghue. 2002. Time discounting and time preference: A critical review. *J. Econom. Literature* 40(2) 351–401.
- Fudenberg, D., D. Levine. 2006. A dual self model of impulse control. *Amer. Econom. Rev.* 96(5) 1449–1476.
- Goldstein, D. G., D. C. Goldstein. 2006. Profiting from the long tail. *Harvard Bus. Rev.* 84(6) 24–28.
- Khan, U. 2007. Where there is a way, is there a will? The effect of future choices on current preferences. *J. Experiment. Psych.: General* 136(2) 277–288.
- Khan, U., R. Dhar, K. Wertenbroch. 2005. A behavioral decision theory perspective on hedonic and utilitarian choice. S. Ratneshwar, D. G. Mick, eds. *Inside Consumption: Frontiers of Research on Consumer Motives, Goals, and Desires*. Routledge, London, 144–165.
- Kirby, K. N. 1997. Bidding on the future: Evidence against normative discounting of delayed rewards. *J. Experiment. Psych.* 126(1) 54–70.
- Kirby, K. N., R. J. Herrnstein. 1995. Preference reversals due to myopic discounting of delayed rewards. *Psych. Sci.* 6(2) 83–89.
- Kirby, K. N., N. N. Marakovic. 1996. Delayed-discounting probabilistic rewards rates decrease as amounts increase. *Psychonomic Bull. Rev.* 3(1) 100–104.
- Laibson, D. 1996. Hyperbolic discount functions, undersaving and savings policy. NBER Working Paper w5635, National Bureau of Economic Research, Cambridge, MA.

- Lichtenstein, S., B. Fischhoff. 1980. Training for calibration. *Organ. Behav. Human Decision Processes* 26(2) 149–171.
- Loewenstein, G. 1996. Out of control: Visceral influences on behavior. *Organ. Behav. Human Decision Processes* 65(3) 272–292.
- Loewenstein, G., D. Prelec. 1992. Anomalies in intertemporal choice: Evidence and an interpretation. *Quart. J. Econom.* 107(2) 573–597.
- Loewenstein, G., R. Thaler. 1989. Anomalies: Intertemporal choice. *J. Econom. Perspectives* 3(4) 181–193.
- Malmendier, U., S. Della Vigna. 2006. Paying not to go to the gym. *Amer. Econom. Rev.* 96(3) 694–719.
- McClure, S. M., D. Laibson, G. Loewenstein, J. D. Cohen. 2004. Separate neural systems value immediate and delayed monetary rewards. *Science* 306(5695) 503–507.
- Milkman, K. L., T. Rogers, M. H. Bazerman. 2008a. Harnessing our inner angels and demons: What we have learned about *want/should* conflicts and how that knowledge can help us reduce short-sighted decision making. *Perspectives Psych. Sci.* 3(4) 324–338.
- Milkman, K. L., T. Rogers, M. H. Bazerman. 2008b. I'll have the ice cream soon and the vegetables later: A study of online grocery purchases and order lead time. HBS Working Paper 07-078, Harvard Business School, Boston.
- Navarick, D. J. 1982. Negative reinforcement and choice in humans. *Learn. Motivation* 13(3) 361–377.
- Nunnally, J. C. 1967. *Psychometric Theory*. McGraw-Hill, New York.
- O'Donoghue, T., M. Rabin. 1999. Doing it now or later. *Amer. Econom. Rev.* 89(1) 103–124.
- Oster, S., F. M. Scott Morton. 2005. Behavioral biases meet the market: The case of magazine subscription prices. *B.E. J. Adv. Econom. Anal. Policy* 5(1) Article 1.
- Phelps, E. S., R. A. Pollak. 1968. On second-best national saving and game-equilibrium growth. *Rev. Econom. Stud.* 35(102) 185–199.
- Phillips, M. 2006. For some Netflix users, red envelopes gather dust. *Wall Street Journal Online* (July 8), http://online.wsj.com/public/article/SB115255814013802582-M1q9pbUxty8N5ItDg5CdaiVP6_A_20070717.html.
- Read, D. 2001. Intrapersonal dilemmas. *Human Relations* 54(8) 1093–1117.
- Read, D., B. van Leeuwen. 1998. Predicting hunger: The effects of appetite and delay on choice. *Organ. Behav. Human Decision Processes* 76(2) 189–205.
- Read, D., G. Loewenstein, S. Kalyanaraman. 1999. Mixing virtue and vice: Combining the immediacy effect and the diversification heuristic. *J. Behav. Decision Making* 12(4) 257–273.
- Rogers, T., M. H. Bazerman. 2008. Future lock-in: Future implementation increases selection of “should” choices. *Organ. Behav. Human Decision Processes* 106(1) 1–20.
- Shrout, P. E., J. L. Fleiss. 1979. Intraclass correlations: Uses in assessing rater reliability. *Psych. Bull.* 86(2) 420–428.
- Solnick, J., C. Kannenberg, D. Eckerman, W. Waller. 1980. An experimental analysis of impulsivity and impulse control in humans. *Learn. Motivation* 11(1) 61–77.
- Strotz, R. H. 1956. Myopia and inconsistency in dynamic utility maximization. *Rev. Econom. Stud.* 23(3) 165–180.
- Thaler, R. H. 1981. Some empirical evidence on dynamic inconsistency. *Econom. Lett.* 8(3) 201–207.
- Thaler, R. H., H. Shefrin. 1981. An economic theory of self control. *J. Political Econom.* 89(1) 392–406.
- Trope, Y., A. Fishbach. 2000. Counteractive self-control in overcoming temptation. *J. Personality Soc. Psych.* 79(4) 493–506.
- Tugend, A. 2006. Getting movies from a store or a mailbox (or just a box). *New York Times* (August 5).
- Wertenbroch, K. 1998. Consumption self-control by rationing purchase quantities of virtue and vice. *Marketing Sci.* 17(4) 317–337.