Contrast Effects in Sequential Decisions: Evidence from Speed Dating*

Saurabh Bhargava
Carnegie Mellon University
sbhar@andrew.cmu.edu

Ray Fisman
Columbia University
rf250@columbia.edu

This version: October 17, 2012

Abstract

We provide an empirical test of contrast effects—a bias where a decision-maker perceives information in contrast to what preceded it—in the quasi-experimental context of "speed dating" decisions. We document that prior partner attractiveness reduces the subsequent likelihood of an affirmative dating decision. This relationship is confined to recent interactions, consistent with a perceptual error, but not learning or the presence of a quota in affirmative responses. The contrast effect is driven almost entirely by male evaluators. Additional evidence documents the effect's linearity with respect to prior partner attractiveness, its amplification for partners of moderate attractiveness, and its partial attenuation with accumulated experience. (JEL C93, D03)

*This paper was shaped by discussions with David Card, Stefano DellaVigna, Botond Koszegi, and Matt Rabin and we owe them special thanks. Dan Black, Raj Chetty, Jack Glaser, Thomas Griffiths, Emir Kamenica, Shachar Kariv, Dacher Keltner, Prasad Krishnamurthy, Robert MacCoun, Ritu Mahajan, Enrico Moretti, Yesim Orhun, Vikram Pathania, Steven Raphael, Jesse Shapiro, Uri Simohnson, Aman Vora, George Wu and seminar participants at the Economics Department at U.C. Berkeley, the Goldman School of Public Policy at U.C. Berkeley, Harvard Business School, the Rady School of Management at UCSD, and the Booth School of Business at the University of Chicago also provided thoughtful comments and feedback. Finally, we thank Sheena Iyengar, Emir Kamenica, and Itamar Simonson for generously providing data.

1 Introduction

In the laboratory, psychologists have documented that sequential evaluations across a variety of social domains are comparative in nature. For example, subjects tasked with sentencing crimes based on written descriptions recommend more lenient sentences if the assignment follows a narrative of a particularly egregious crime. (Pepitone and DiNubile 1976).[1] This transient contrastive influence of recent context on subsequent perceptions is known as a *contrast effect*.

Such patterns of sequential decisions could also be the result of learning or the presence of a budget constraint in permissible responses, (i.e., a "quota"), rather than evidence of an error in perception. In the opening example, learning of an egregious crime may cause a subject to "rationally" update her beliefs regarding criminality and to judge a subsequent crime less punitively on a relative scale. Distinguishing among these explanations is of theoretical import in building models of individual decision-making, and also of practical significance in understanding the welfare consequences of weighting recent experiences "too much." Contrast effects potentially influence outcomes in a wide range of important sequential decisions including employee hiring, judicial sentencing, the evaluation of investments, and medical diagnoses.

This paper tests for the existence of contrast effects in a unique field setting—speed dating—where we are able to distinguish perceptual errors from other plausible explanations. Speed dating refers to an organized match-making event in which men and women sequentially meet potential matches through a series of short interactions, or dates, each lasting a few minutes. At the close of each interaction, subjects (hereafter, "evaluators") are instructed to privately record assessments of partner (hereafter, "target") attributes, as well as a "yes/no" declaration of romantic interest.[2] In the event of mutual interest, organizers distribute contact information the following day to both parties. In this context, we test for contrast effects by examining whether perceptions of prior target attributes, such as physical attractiveness, temporarily distort subsequent romantic decisions. Our analysis is based on data from 16 speed dating sessions organized in the New York area from 2002 to 2004 (Fisman et al. 2006; Fisman et al. 2008). Nearly 500 participants—typically graduate or professional school students in their 20s—collectively made over 7,000 romantic decisions.

[1] Similar effects are documented in several other studies. Damisch, Mussweiler and Plessner (2006) found that experienced judges of gymnastic competitions evaluate videos of a routine less favorably if preceded by a higher quality routine. In a second example, subjects exposed to an advertisement with an attractive female model judge subsequent yearbook photographs as less attractive than those not exposed to the model (Kenrick and Gutierres 1980).

[2] Throughout the paper, we refer to decision-makers as "evaluators", and the targets of their decisions as "targets." In a given interaction, of course, a participant is both an evaluator and a target.

Speed dating is a compelling setting from which to identify contrast effects in the field. First, we can estimate "objective" attribute quality through ratings of third party research assistants. Second, in the absence of explicit random assignment, we use these objective measures to establish that the order of evaluation is effectively random. Third, our large, repeated decision, within-subject samples help to rule out alternative explanations, such as learning, or quotas for affirmative decisions. Finally, we can infer preferences revealed from consequential "yes/no" dating decisions rather than numerical assessments where respondents face no truth-telling incentives and where responses may be subject to biases due to rescaling. Abstracting from these methodological considerations, we also believe that this setting is of per se interest. A large literature in economics focuses on sorting and efficiency in matching markets including a growing sub-literature that specifically targets romantic matching (e.g., Becker 1973; Fisman et al. 2006; Hitsch et al. 2010). Our study investigates a systematic and potentially distortionary behavioral bias in the context of an important matching market.

We present two main results which implicate contrast effects in dating choices. Our primary finding is that an evaluator's dating decision is negatively correlated with the attractiveness of the prior target even after controlling for current target attributes and the evaluator's own selectiveness. A one unit rise in prior target attractiveness on a 1-10 scale produces a 1.9 percentage point drop in an evaluator's willingness to date the current target. This effect is 18 percent as large as the positive influence of an equivalent change in the current target's attractiveness, and is almost entirely driven by male evaluators for whom the influence of a recent target is 31 percent as large as that of a current target. We further find that an evaluator's subjective ratings of attractiveness are also negatively influenced by past target attractiveness suggesting that the observed effect is mediated through perceptions of attractiveness. Placebo tests on lead—as opposed to lagged—targets, and simulated data with arbitrarily reordered dater sequences, confirms the idiosyncratic influence of prior target attractiveness.

While the negative correlation in judgments across attractiveness is consistent with contrast effects, there are two main alternative explanations—evaluator learning and the presence of quotas for high or low evaluations. For example, a dater who updates ex-ante beliefs regarding the distribution of target beauty will act less generously after encountering an earlier, very attractive target. Further, a dater with limited time, money or attention, might limit himself to a small number of affirmative responses which would also create a negative correlation in observed decisions.

To distinguish contrast effects from these two alternatives, our empirical identification appeals to predictions of a simple framework of sequential decision-making detailed in Bhargava (2012). This model of signal extraction features a Bayesian decision-maker who makes choices based on the perceived quality of a series of targets, inferences regarding the relative standing of each target, and preferences over the accuracy of assessments. The model additionally considers constraints due to evaluative quotas as a finite dynamic programming problem. The model generates two intuitive predictions to differentiate between "standard" decision-makers and "behavioral" ones subject to contrast effects. First, contrast effects should result in a negative correlation across evaluations, as we find in the data. Second, for the evaluator subject to contrast effects, the negative correlation should be more pronounced for recent as compared to more distant evaluations. The differential influence of recent and distant past is caused by the transient influence of a contrast effect on subsequent perceptions. A Bayesian unaffected by perceptual distortions should not be sensitive to the order in which evaluations are assigned even if constrained by a quota for high or low evaluations.

Consistent with this prediction, the second main finding of our analysis is that the contrastive influence of prior targets is confined to the recent past. Estimating current dating decisions as a function of lagged target attractiveness across three past "dates", we find that only the first lag is significant and can reject the equality of recent and more distant lagged targets. We consider and discuss alternative explanations, beyond learning or quotas in affirmative decisions, such as limited memory, base-rate neglect, or some combination thereof. We argue that the sharpness with which the contrastive influence of past targets fades is evidence against plausible explanations involving only limited memory.

In a series of extensions, we examine the factors that moderate the strength of contrast effects. First, we find that the contrast effect appears linear with respect to recent targets ranked by attractiveness quintiles. Second, we find that a pair of highly attractive targets triggers a larger contrast effect than a single target. Third, as presaged by psychological research, we find that contrast effects are heightened in the presence of current targets who are of moderate attractiveness. Finally, using two measures of experience—the accumulation of dates within each session and self-reported past dating experience—we find that experience attenuates, but does not eliminate, the magnitude of the contrast effect.

Our research contributes to the body of work that studies how comparative assessments shape evaluations and more broadly to the literature on the influence of context on economic decisions. This includes marketing research describing departures from the standard rational choice framework to explain product choice and price perception (e.g., Simonson and Tversky 1992; Tversky and Simonson 1993), as well as research which provides explanations for context-based effects using rational inference (e.g., Kamenica 2008). Our work also relates to research suggesting perceptual or inferential biases in repeated decisions, including research on quasi-Bayesian updating (e.g., Tversky and Kahneman 1971; Rabin and Schrag 1999; Rabin 2002) and categorical biases (e.g., Quattrone and Jones 1980; Fryer and Jackson 2008). More specifically, while there is a rich literature exploring contrast effects in sequential decisions in the laboratory, this paper is part of a small empirical literature that has begun to investigate the role of comparisons in field settings, including work on housing and commuting choices (Simonsohn and Loewenstein 2006; Simonsohn 2006) and willingness to pay for art (Beggs and Graddy 2009). The paper most closely linked to the present study, and the only other work involving social evaluations in the field, shows that leniency in judicial sentencing decisions is influenced by exposure to extreme recent crimes (Bhargava 2012). Finally, the difference we document across gender contributes to an emerging literature that connects decision-making biases to individual attributes (e.g., Benjamin et al. forthcoming; Choi et al. 2011; Stanovich and West 1998).

Our paper differs from earlier research on contrast effects in that we exploits a unique, quasi-experimental, high-frequency setting in the field that permits us to differentiate perceptual errors due to contrast effects from alternative explanations such as learning and evaluative quotas. Moreover, the paper addresses a long-standing ambiguity that characterizes most observed effects in the laboratory. One can interpret most laboratory findings as either evidence of an actual change in perceptual experience or a rescaling of the numerical response variable (e.g., Scherer and Lambert 2009). Our analysis, by relying on actual decisions rather than numerical assessments, within-subject estimates, and exogenous assessments of target values, does not suffer from this interpretive ambiguity. Overall, we find contrast effects that are comparable in magnitude to analogous studies in the laboratory. Finally, our setting permits us to demonstrate that the psychological bias endures even after repeated decisions.

2 Data and Experimental Design

Our experimental design relies on speed dating sessions in which participants engage in a series of short "dates" to identify romantic compatibility across a large pool of potential mates (Fisman et al. 2006; 2008). Other researchers have used speed dating to study the determinants of romantic selection and attraction (e.g., Kurzban and Weeden 2005; Belot and Francesconi 2006; Fisman et al. 2006), racial preferences in dating (Fisman et al. 2008), as well as differences in stated and revealed romantic preferences (Eastwick and Finkel 2008).

The advantage of this research design is that it allows us to observe sequential decisions and infer preferences in a setting similar to what one might expect in the "real world" while allowing for some experimental control. Since the first speed dating events were organized in 1998, several private firms have popularized the format across the United States. In order to maintain the realism of the experimental setting, the script for all events is based on a modified version of that used by *HurryDate*, a commercial firm which was the largest organizer of speed dating events in New York at the time our experiments took place.

Experimental Procedure and Setting. Our data comprises 16 speed dating sessions organized in the New York area from 2002 to 2004 by Fisman et al. (2006; 2008).[3] Participants for speed dating sessions were recruited from the campus of Columbia University and were, for the most part, students enrolled in a graduate or professional school. Sessions were held in a closed room of a local bar/restaurant during weekday evenings.[4] The aesthetic details of each event—table arrangements, lighting, music—were fixed across days. The notable experimental difference across sessions was group size which varied from 18 to 44 participants. A total of 474 participants collectively made 7684 decisions.

After arrival and registration, participants were handed a name-tag, clipboard, scorecard, and assigned an anonymous ID. The scorecards were designed so that after each date, daters could record a "yes/no" declaration of romantic interest on a line labeled "Decision", and rate their target on a 1-10 scale across six target specific attributes: Ambition, Attractiveness, Fun, Intelligence, Shared Interests and Sincerity.[5] Hosts then directed the men and women to seat themselves on opposite sides of adjacent two-person tables. Importantly, during the course of the evening, two research assistants (RAs) independently evaluated each participant's objective attractiveness on a 1-10 scale.[6]

Each round, daters interacted for four minutes and were then given one minute to (privately) appraise their partners. In accordance with *HurryDate* norms, males then shifted to the adjacent table and the dates continued until each male had been paired with each female. The sequence with which each evaluator interacted with targets was thus fixed across evaluators within a session with the exception of staggered sequence starts. In the event of mutual interest, organizers later distributed contact information to both members of the pair. A more detailed description of the experimental setting and procedure is provided in Fisman et al. (2006).

Our main analysis relies on two variables. The first is the "yes/no" decision made by each evaluator, e, with respect to target, t, which we denote by $Dec_{e,t}$. This indicator

[3] Originally, 21 sessions were organized. Data from 5 of these sessions were eliminated due to various procedural problems (see Fisman et al. 2006).

[4] Generally two sessions were scheduled for a given evening. Participants were randomly assigned to one of the two sessions.

[5] Specifically, with respect to the "yes/no" decision, the scorecard reads "Decision" and instructs the subject to circle either "yes" or "no" for each partner.

[6] The RAs were instructed to provide ratings as though they were judging a beauty contest (and hence rating participants on consensus views of beauty rather than their own idiosyncratic preferences). The ratings of the two RAs were highly correlated within each session (ρ = .70), and such RA ratings of attractiveness are a norm in social psychology. It is conceivable that the RAs themselves may exhibit contrast effects in their evaluations, despite knowingly serving as "objective" raters for the study. However, given that we combine ratings of RAs who evaluate targets in an unknown order, any bias in RA evaluations should simply add noise to our estimates.

variable allows us to infer target preferences (assuming no strategic behavior) and serves as the dependent variable for most of the analysis. The second variable of interest, $Attract_{e,t}$, is the mean "objective value" for target attractiveness as scored by the research assistants.

Random Ordering of Targets. One assumption of the research design which underlies much of the subsequent analysis—particularly the examination of alternative explanations—is that the order that participants were seated is either random or is random conditional on observable attributes. Imagine some component of desirability, (e.g., confidence), is unobserved and not perfectly correlated with our measure of attractiveness. If daters' ordering is negatively correlated with respect to such attributes, one could identify a spurious negative correlation between current dating decisions and past target attractiveness even after controlling for current target attractiveness. It is difficult to imagine how such negative correlation might come about, particularly given the logistical and procedural details of each evening. One possibility is if some participants are aware of the presence of contrast effects and such awareness is correlated with (unobserved) desirability and prompts strategic seating. A second problematic scenario could arise if some component of undesirability, (e.g., self-absorption), is unobserved, and not perfectly correlated with our measure of attractiveness. In this case, if the order of daters is positively correlated with this attribute, one would again identify a spurious negative correlation between current decisions and past target attractiveness, even after adding controls.[7]

One test for random ordering—at least to the extent that it is correlated with observables—is to examine whether the observable attributes of a dater are correlated with the observable attributes of a preceding dater. As an initial implementation of this test, we measure the correlation between lagged and current dater attractiveness. While intuition might suggest estimating this autocorrelation using a panel regression, such an estimate would be biased due to the considerable session specific heterogeneity in the ratings of dater attractiveness. This heterogeneity may be due both to variation in RA measurement as well as real differences across the populations which attend each session. The usual solution to this omitted variable would be to include session fixed effects so that we estimate a regression of the following form:

$$Attract_d = \alpha + \gamma Attract_{d-1} + \eta_j + \varepsilon_{d,j} \tag{1}$$

where $Attract_d$ refers to the attractiveness of dater, d, and session, j, specific variation is indicated by fixed effects, η_j.

As first documented by Nickell (1981), the panel estimation above, with its lagged

[7] Note that the outlined scenario also implies that decisions to date are positively correlated within a session. We test for this via simulations, unreported here, and do not find any evidence for positive correlations in such decisions.

dependent variable, and session fixed effects, produces an inconsistent and downward biased estimate of the lagged dependent coefficient, $\widehat{\gamma}$ (1981). This attenuation is tied to the time-length of the panel and may, in our data, be particularly pronounced (Phillips and Sul 2007).

We employ a non-parametric strategy to overcome this bias. We first generate a set of simulated data by arbitrarily reordering daters within a given session and then estimating the above model with the simulated dataset to produce a bootstrap coefficient estimate. We repeat this procedure to generate a sampling distribution of the coefficient estimate of interest. Note that the estimates rely on actual data on dater attractiveness—it is only the ordering of targets within a session that is randomly regenerated. Finally, we locate the coefficient estimate from authentically ordered data within the distribution of bootstrap estimates. This comparison yields a percentile rank that we can interpret as the likelihood of rejecting the null hypothesis of random ordering through sampling variation alone. We simulate 10,000 such regressions at the dater, rather than the interaction, level for computational ease.[8]

We perform the described exercise for all daters, and then for each gender separately. The first row of Table 1 reports the estimated coefficient from Equation (1) using actual data, as well as the empirical "p-values" from the comparison with coefficients from simulated data. For example, the first row indicates that 31% of the simulated coefficients were less than the -.043 point estimate generated from the authentically ordered data. The estimates, reported in the first row, provide no evidence to suggest that male or female dater order is correlated with respect to physical attractiveness.[9]

Since research assistants measure participant attractiveness, its analysis constitutes the most convincing test of random ordering. We can, however, test for randomness based on other attributes for which we have plausibly unbiased ratings. Because dater order is fixed, subjective evaluator assessments of dater attributes, described above, may be sensitive to systematic bias due to contrast effects. However, each evaluator's rating of the first target he or she encounters should be unaffected by any contrast to future targets. Under this assumption, we can treat the first rating received by a target across the six attributes, including attractiveness, as the "objective," or unbiased, basis for additional tests of random order.

[8] Our simulations suggest that we achieve convergence in empirical p-values with less than 500 iterations.

[9] An alternative test of random dater order, suggested by an anonymous referee, is to estimate the correlation of the attractiveness ratings after standardizing such ratings by session. We perform this exercise and find that for female targets, the correlation between adjacent daters is -0.02 (p= 0.76), and for male targets, the correlation is -0.05 (p= 0.37).

Table 1

NON-PARAMETRIC TEST OF RANDOM DATER ORDER

	LAGGED ATTRIBUTE COEFFICIENT WITH EMPIRICAL P-VALUES		
	Dater Population		
	All (1)	Male (2)	Female (3)
Attractiveness (Research Assistants)	-0.043 p = .31	-0.136 p = .18	-0.059 p = .60
Ambition	0.034 p = .48	-0.087 p = .47	-0.042 p = .66
Attractiveness	-0.036 p = .41	-0.094 p = .39	-0.076 p = .47
Fun	0.036 p = .83	-0.043 p = .67	-0.015 p = .79
Intelligence	0.018 p = .77	-0.088 p = .41	0.025 p = .92
Shared Traits	-0.057 p = .43	-0.113 p = .34	-0.079 p = .53
Sincerity	-0.004 p = .81	0.011 p = .91	-0.090 p = .39

Notes: Table 1 reports results from tests of random dater ordering. The first row reports the lagged coefficient estimates generated from Equation 1 which is described in the text. The dependent variable for the first row estimates is the RA assessment of target attractiveness. The remaining rows report lagged coefficient estimates from Equation 1 modified such that the dependent variable is the evaluator's assessment of the indicated target attribute elicited in the first dating round. All regressions are at the date, rather than interaction, level. The empirical p-values communicate the position of the estimated lagged coefficients in the distribution of coefficient estimates from regressions on simulated data. The data is simulated by iteratively reshuffling partner order within a session in a manner described in the text. For example, the p = .31, in the first cell, indicates that 31% of the simulated coefficients are smaller than the coefficient estimated from authentically ordered data (-.043). The columns report estimates of the pooled sample, males, and females, respectively.

The remaining rows of Table 1 report the lagged coefficient estimates for each of the six attributes, including attractiveness, using the actual data, as well as the p-values generated by comparison with simulated estimates. Overall, the table offers no systematic evidence that daters, either male or female, are non-randomly ordered across observable attributes. This result is consistent with the observations of on-site event organizers.

3 Empirical Analysis

3.1 Identification Strategy

The empirical identification of a contrast effect relies on straightforward intuition presented in Bhargava (2012). Consider a framework which features a Bayesian decision-maker who evaluates a sequence of targets based on her perception of the targets' intrinsic qualities and her preference over the accuracy of her evaluations. Evaluations are made on a continuous scale, but can be translated without loss of generality to a binary "yes/no" decision.

In this framework, we can decompose the intrinsic quality of each target into two components such that: $q = s + \psi$. The first, s, is a systematic component which is common across all targets and drawn from some normal distribution, while the second, ψ, is an idiosyncratic component unique to each target, i.i.d., and also drawn from some normal distribution. As a decision-maker proceeds through rounds, she learns the value of the systematic component of quality through Bayes' Rule. The model assumes that a relative decision rule governs evaluations.

In this context, imagine a speed dater who must evaluate a sequence of romantic targets and decide whom to date.[10] An evaluation in period t occurs in three steps: The dater perceives a signal of quality (q_t), infers the idiosyncratic component of dater quality ($\widehat{\psi}_t = q_t - \hat{s}_t$, where $\hat{s}_t = E_t(s \mid q_1, ..., q_t)$), and then maps the inferred quality to a final decision (dec_t) based on a utility function which captures a preference for accuracy. For simplicity, preferences are specified as the minimization of the sum of least-square errors between inferred quality and final evaluations each period: $-\sum_{j=1}^{t}(dec_j - \psi_j)^2$ so that the decision-maker sets $dec_t^* = \widehat{\psi}_t$.

The first of two main empirical predictions is that, for any given dater, evaluations should be negatively correlated across periods such that $\partial dec_{t-k}/\partial q_{t-k-1} \leq 0$ for all k. The negative correlation emerges from three possible mechanisms. The first is learning. If the dater encounters an attractive target in one period, she will update her priors on the underlying distribution of target beauty ($\partial \hat{s}_t/\partial q_{t-1} > 0$), and will judge a subsequent target more punitively ($\partial \widehat{\psi}_t/\partial \hat{s}_t < 0$) such that, at least in early rounds, a high evaluation in one period will result in a lower expected evaluation in the subsequent period.

A second explanation for the negative correlation is if the dater is subject to a quota limiting the number of affirmative evaluations she is able to assign. The intuition for the effect of the quota on behavior can be explained with a simple decomposition. Suppose a dater in a particular period awards a "yes". If the evaluation binds the quota constraint, the dater will reject subsequent targets. If the decision does not bind the constraint, the dater assigns a "yes" only if target quality is far enough above a relative threshold to compensate for the lost option value of assigning future high evaluations, less any penalty incurred in the current period due to inaccuracy. This functional threshold above which target quality must reach in order for the dater to award a "yes" in a given period is a positive and monotonic function of the number of previously assigned high evaluations. If in the last period the dater assigns a "yes", the subsequent functional threshold rises, and

[10] The framework supposes that the decision-maker is not allowed to revisit evaluations once they have been rendered. This assumption is consistent with the speed dating paradigm.

the dater, all else equal, will be more punitive in evaluation of future targets.

Finally, the negative correlation may arise through a contrast effect. A decision-maker subject to contrast effects perceives quality q_t^c not only as a positive function of the quality of the current target, but as a negative function of the quality of past period targets. Importantly, the decision-maker is unaware of this error and updates as if her perception were accurate. A dater who encounters an attractive target in one period misperceives the subsequent target ($\partial q_t^c / q_{t-1}^c < 0$), and delivers a lower subsequent evaluation ($\partial dec_t^c / \partial q_{t-1}^c \leq 0$).

A second prediction allows one to differentiate between "rational" behavior such as learning or adherence to a quota constraint, and behavior consistent with a perceptual contrast effect. For the former, the influence of the quality of a past target on the evaluation of a subsequent one is not a function of the distance between them (i.e., $\partial dec_t / \partial q_{t-k} = \partial dec_t / \partial q_{t-l}$ for all $k,\ l$). However, for decision-makers subject to contrast effects, the influence of the quality of a target on a subsequent evaluation is negatively related to the distance between the targets such that $\left| \partial dec_t^c / \partial q_{t-k}^c \right| \geq \left| \partial dec_t^c / \partial q_{t-l}^c \right|$ for all $0 < k < \ l$.

The basic intuition for this prediction comes from the principle of exchangeability which holds that for i.i.d. sequences of random variables, the joint probability of any pair of realizations is invariant to permutation (Kreps 1988). A consequence of this property is that for a standard decision-maker, the order of realizations should be irrelevant. A Bayesian free from quotas should treat all past observations equally (i.e., $\partial \hat{s}_t / \partial q_{t-k} = \partial \hat{s}_t / \partial q_{t-l}$ for all k,l), and, as long as costs are not convex, a decision-maker subject to a quota constraint should be sensitive to the number of high evaluations already handed out, but not the order in which they occur. For the decision-maker subject to a temporary contrast effect, the recent past is more influential than the more distant past. That is, an attractive target encountered in period 4 should exert less influence on a period 9 decision than a comparably attractive target encountered in period 8.

3.2 Evidence from Speed Dating

Recent Target Attractiveness and Evaluator Decisions. We begin our analysis by examining the influence of the attractiveness of a prior target on an evaluator's current dating decision. A dynamic panel specification with one lag formally tests for the relationship between a current decision and the attractiveness of the prior target:

$$Dec_{e,t} = \alpha + \gamma Attract_{e,t} + \lambda Attract_{e,t-1} + \xi_e + \varepsilon_{e,t} \tag{2}$$

In this specification, λ captures the influence of lagged target attractiveness, $Attract_{e,t-1}$, on a current decision, $Dec_{e,t}$, after controlling for current target attractiveness, $Attract_{e,t}$, and fixed effects ξ_e to account for evaluator specific variation. λ is identified in this model if there is sequential exogeneity conditional on ξ_e:

$$E(\varepsilon_{e,t}|\ Attract_{e,1}, Attract_{e,2}..., Attract_{e,t}, \xi_e) = 0$$

The contemporaneous error here is uncorrelated with past or present covariates. If dater order is conditionally random, then the above assumption of exogeneity is satisfied. Errors are clustered by target to account for the fixed sequence with which evaluators encounter targets.

Consistent with a contrast effect, the first column in Table 2 implies that a one unit rise in prior target attractiveness leads to a 1.9 percentage point drop in current willingness to date. This is relative to an overall willingness to date of 42 percent. The contrastive influence of recent target attractiveness is 18 percent as large as the positive influence of an equivalent one unit change in current target attractiveness.

We observe a sharp gender asymmetry. While both male and female dating decisions are determined by contemporaneous target attractiveness, only male evaluators are sensitive to prior target attractiveness. For males, the contrastive influence of recent target attractiveness is 31 percent as large as the influence of current target attractiveness.

If contrast effects are responsible for the observed negative correlation, the correlation should exist with respect to past, but not future, targets. As a placebo check, the final three columns of Table 2 estimate an analogous specification that tests for the influence of future targets on current dating decisions, after controlling for current target attractiveness and evaluator fixed effects.[11] One can also interpret the placebo test as an additional, indirect check for the random ordering of participants. The table provides no evidence that future target attractiveness negatively influences current dating decisions.

As an additional placebo check, we compare estimates from the actual data to estimates derived from data simulated in a manner that parallels the test of random ordering. That is, we randomly reshuffle target order, this time at the level of the evaluator, and then estimate the influence of prior targets on current decisions, after controlling for current target attractiveness and evaluator fixed effects. This exercise produces a sampling distribution of bootstrap estimates from which we can calculate the empirical p-value of the null hypothesis that lagged target attributes have no influence on current decisions. Again, we estimate 10,000 regressions for each simulation. The lower panel of Table 2 reports p-values from this exercise. The simulations confirm that the authentically ordered data produces coefficient values larger in absolute magnitude than those from random dater ordering and corroborate the previously exhibited gender asymmetry.

[11] Our test assumes that evaluators do not attend sufficiently to future targets for such targets to also trigger a contrast effect. Given the various distractions that characterize the speed dating experience, our belief is that it is likely difficult to attend to anyone other than one's current dating partner.

Table 2

EFFECT OF RECENT TARGET ATTRACTIVENESS ON DATING DECISIONS

	DEPENDENT VARIABLE - DECISION TO DATE (OLS)					
	BASELINE Target Population			PLACEBO Target Population		
	All (1)	Male (2)	Female (3)	All (4)	Male (5)	Female (6)
Attractiveness	0.107*** (0.008)	0.112*** (0.012)	0.103*** (0.010)	0.113*** (0.008)	0.119*** (0.012)	0.107*** (0.010)
Attractiveness - Lag	-0.019** (0.008)	-0.035*** (0.011)	-0.004 (0.010)			
Attractiveness - Lead				0.012 (0.008)	0.013 (0.012)	0.011 (0.010)
N	N = 7200	N = 3600	N = 3600	N = 7200	N = 3600	N = 3600
R^2	0.33	0.34	0.31	0.33	0.33	0.30
Empirical P-Value for Attractiveness - Lag 1	p < .01	p < .01	p = .28			

Notes: Table 2 reports results from tests of the influence of recent target attractiveness on current dating decisions. The dependent variable is a binary variable indicating a subject's "yes" / "no" decision to date each round. Target attractiveness is an average rating from two research assistants on a scale from 1-10. Fixed effects control for evaluator specific decisions across all specifications. The first three columns report results of Equation 2 for the sample of pooled, male, and female targets, respectively, while the final three columns report analogous results for a placebo test of lead target influence. Regressions are weighted to account for the varying number of targets encountered by each dater. Standard errors are robust and clustered at the target level and are reported parenthetically. The bottom panel reports empirical p-values for the lagged coefficients from the estimation of Equation 2. The p-values indicate the likelihood of obtaining a coefficient estimate, from simulated data, smaller than the one produced from authetically ordered data.

* significant at 10%; ** significant at 5%; *** significant at 1%

Recent Target Attractiveness and Evaluator Ratings. In order to confirm that the observed negative correlation operates through distorted perceptions of attractiveness, we examine the influence of recent target attractiveness on subjective evaluator ratings of current target attractiveness. (It is worth noting that these results should be interpreted with some caution, owing to the use of stated rather than revealed preferences, and the possible rescaling of the response variable over time.)

Table 3 estimates Equation 2 after substituting decisions with evaluator ratings of target attractiveness as the dependent variable. The coefficient on prior target attractiveness is negative, though not significant, for the pooled sample of male and female evaluators. However, for male evaluators there is a negative influence of past attractiveness on current ratings, significant at the 5 percent level. The relative influence of past, as compared to current, target attractiveness on ratings is smaller than it is on decisions (approximately 15 percent, relative to the 31 percent effect on past decisions). This difference may be due to a nonlinear mapping between ratings and decisions. Ratings of female evaluators appear insensitive to past attractiveness.

As a placebo test of the proposed mechanism, the last set of columns demonstrates that future target attractiveness does not influence current ratings. Additionally, the lower panel of the table reports the empirical p-values from a comparison of the coefficient estimates to the sampling distribution of bootstrap estimates produced from 10,000 regressions on simulated data. The simulation confirms the prior results.

Table 3

EFFECT OF RECENT TARGET ATTRACTIVENESS ON SUBJECTIVE RATINGS

	DEPENDENT VARIABLE - SUBJECTIVE ATTRACTIVENESS RATING (OLS)					
	BASELINE Target Population			PLACEBO Subject Population		
	All (1)	Male (2)	Female (3)	All (4)	Male (5)	Female (6)
Attractiveness	0.672*** (0.035)	0.651*** (0.048)	0.693*** (0.049)	0.686*** (0.035)	0.666*** (0.049)	0.704*** (0.049)
Attractiveness - Lag	-0.030 (0.034)	-0.099** (0.040)	0.036 (0.053)			
Attractiveness - Lead				0.034 (0.034)	0.044 (0.045)	0.025 (0.051)
N	N = 7005	N = 3503	N = 3502	N = 7023	N = 3511	N = 3512
R^2	0.47	0.47	0.46	0.48	0.46	0.47
Empirical P-Values for Target Attractiveness - Lag 1	p = .08	p < .01	p = .91			

Notes: Table 3 reports results from tests of the influence of recent target attractiveness on evaluator assessments of attractiveness. The dependent variable is a variable ranging from 1-10 which indicates the assessment of target attractiveness each round. Target attractiveness is an average rating from two research assistants on a scale from 1-10. Fixed effects control for evaluator specific decisions across all specifications. The first three columns report estimation results of a modified version of Equation 2 for the sample of pooled, male, and female targets, respectively, while the final three columns report analogous results for a placebo test of lead target influence. Regressions are weighted to account for the varying number of targets encountered by each dater. Standard errors are robust and clustered at the target level and are reported parenthetically. The bottom panel reports empirical p-values for the lagged coefficients from the estimation of Equation 2. The p-values indicate the likelihood of obtaining a coefficient estimate, from simulated data, smaller than the one produced from autheticaly ordered data.

* significant at 10%; ** significant at 5%; *** significant at 1%

Contrast Effects and Overall Dating Success. An alternative way to characterize the magnitudes of these effects is to estimate the change in the overall number of "yes" responses a dater might earn in the absence of contrast effects. In this context, such a calibration is relevant given that a target will always be evaluated in a fixed sequence with the exception of the initial round.

A dater-level regression tests the relationship between the total number of "yes" evaluations received by a dater on that dater's attractiveness, the attractiveness of the prior dater in the sequence, as well as fixed effects to control for session specific variation: $Yes_{d,j} = \alpha + \gamma Attract_{d,j} + \lambda Attract_{d-1,j} + \xi_d + \varepsilon_{d,j}$. The estimation indicates that a one unit rise in current target attractiveness results in 1.9 additional "yes" responses on average. A one unit rise in prior dater attractiveness leads to a 0.25 decrease ($p < 0.10$)

in such responses. This effect is entirely driven by female targets, (i.e., male decision-makers), for whom a unit change in prior attractiveness yields a 0.49 decrease ($p < 0.02$) in the number of affirmative responses.

What might this mean for a particularly fortunate, or unfortunate, dater? Given that the median number of affirmative responses received by a female target is 8, a change in prior target attractiveness of 3 units—roughly equivalent to a movement from the 25th to the 75th percentile in the attractiveness distribution—would drop overall "yes" responses by 1.5, or 19%. With such a fall in yield, the female dater would roughly move from the 50th to the 40th percentile in apparent desirability (approximately equivalent to a 1 rating point drop on a scale from 1 to 10).

3.3 Alternative Explanations

Learning and Quotas. While the main finding of a negative correlation between target attractiveness and subsequent decisions is consistent with a contrast effect, it may also be reconciled with explanations based on learning or the presence of a quota in the number of affirmative responses. In an effort to differentiate contrast effects from these two alternatives, we first test whether the influence of a target on a future decision decays as the intervening distance between the decisions increases.

We formally compare the effect of recent and distant past targets on current decisions by estimating the following model which includes first, second and third lagged covariates of attractiveness:

$$Dec_{e,t} = \alpha + \gamma Attract_{e,t} + \lambda_1 Attract_{e,t-1} + \lambda_2 Attract_{e,t-2} + \lambda_3 Attract_{e,t-3} + \xi_e + \varepsilon_{e,t} \quad (3)$$

The results of the estimation, summarized in the first three columns of Table 4, indicate that contrast effects decay sharply. For the pooled sample and the sample restricted to males, only the first lagged covariate of target attractiveness is negative and statistically significant. An F-test rejects the null that the autocorrelation between the current and first lagged period is equal to the autocorrelation between the first and second lagged period ($p < 0.05$) or is equal to the autocorrelation between the first and third lagged period ($p < 0.10$). While contrast effects might plausibly produce the one period decay implied by the table, such rapid decay is harder to reconcile with explanations based on learning or quota constraints.

Table 4

EFFECT OF RECENT VERSUS DISTANT PAST AND EXPLICIT CONTROLS FOR PAST RESPONSES ON DATING DECISIONS

	DEPENDENT VARIABLE - DECISION TO DATE (OLS)					
	RECENT VERSUS DISTANT PAST Target Population			EXPLICIT CONTROLS FOR QUOTA Target Population		
	All	Male	Female	All	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Attractiveness	0.109***	0.110***	0.108***	0.108***	0.110***	0.107***
	(0.008)	(0.012)	(0.010)	(0.008)	(0.012)	(0.010)
Attractiveness - Lag 1	-0.021***	-0.031***	-0.011	-0.017**	-0.028***	-0.007
	(0.008)	(0.010)	(0.011)	(0.008)	(0.010)	(0.011)
Attractiveness - Lag 2	-0.008	0.002	-0.018	-0.006	0.004	-0.014
	(0.008)	(0.011)	(0.011)	(0.008)	(0.011)	(0.011)
Attractiveness - Lag 3	-0.004	-0.004	-0.003	-0.002	-0.003	0.000
	(0.008)	(0.011)	(0.011)	(0.008)	(0.011)	(0.011)
Fixed Effects for Past Number of Yes's				X	X	X
N	N = 6252	N = 3126	N = 3126	N = 6252	N = 3126	N = 3126
R^2	0.35	0.36	0.32	0.36	0.37	0.34
Exponential Decay Parameter	0.186***	0.201***	0.082			
	(0.044)	(0.060)	(0.067)			

Notes: Table 4 reports results from tests of the influence of recent, and more distant, target attractiveness on current dating decisions. The dependent variable is a binary variable indicating a subject's "yes" / "no" decision to date each round. Target attractiveness is an average rating from two research assistants on a scale from 1-10. Fixed effects control for evaluator specific decisions across all specifications. The first three columns report results of Equation 3 for the sample of pooled, male, and female targets, respectively, while the final three columns report analogous results for the same test but after controlling flexibly for the number of prior affirmative decisions made by an evaluator. Regressions are weighted to account for the varying number of targets encountered by each dater. Standard errors are robust and clustered at the target level and are reported parenthetically. The bottom panel reports estimates of the non-linear model of recent target influence on current dating decisions described in the text.

* significant at 10%; ** significant at 5%; *** significant at 1%

As an additional test of the quota hypothesis, we estimate the same model but include an explicit, and flexible, control for an evaluator's history of responses. The last three columns of Table 4 indicate that the negative correlation between target attractiveness and subsequent decisions persists even after including fixed effects to control for the number of past affirmative responses.

It is possible that the influence of target attributes on current decisions decays non-linearly. One candidate functional form comes from assuming an exponential decay in the linear relationship between decisions and target attractiveness across periods. For the following model, we use a maximum likelihood estimation:

$$Dec_{e,t} = \alpha + \gamma Attract_{e,t} - [\beta * \gamma Attract_{e,t-1} + \beta^2 * \gamma Attract_{e,t-2} + \beta^3 * \gamma Attract_{e,t-3}] + \xi_e + \varepsilon_{e,t}$$

The parameter β denotes the factor of exponential decay. The estimates, reported near the bottom of Table 4, show a statistically significant decay of 81% (i.e., 1 - 0.19) in target attractiveness across periods (and a decay of 80% for male evaluators). That is, relative to the influence of target attractiveness in the current period, the influence of the most recent target is 19% as large, while the influence of the target two periods in the past is 4% as large, and so forth. The rapid rate of decay implied by these estimates corroborate earlier findings—recent, but not distant, prior targets exert negative influence on dating decisions, and only male evaluators exhibit a statistically significant contrast effect.

Additional Alternative Explanations. An unexplored possibility, beyond learning or the presence of quotas for high or low evaluations, is that the effects are driven by limited memory. For example, in the extreme instance where a decision-maker only recalls the last target, Bayesian learning would prompt a low evaluation in a period subsequent to every naturally occurring high evaluation. This would explain both a negative correlation in evaluations as well as the differential influence of recent as compared to distant targets.

However, for a truly "rational" Bayesian, the presence of limited memory alone is unlikely to produce these results. A rational decision-maker, even one saddled with limited memory, should still update optimally so long as she is able to commit a small number of sufficient statistics—such as the empirical mean and sample size—to memory. It is not obvious that monitoring such sufficient statistics is substantially more burdensome than tracking the attractiveness of the most recent target.

A decision-maker subject to limited memory coupled with an additional bias such as selective recall, a counting heuristic, or base-rate neglect might behave in a manner indistinguishable from that induced by a contrast effect. With such a heuristic, the single period decay evidenced in the data implies that to implicate limited memory, a dater's memory would last only about 5 minutes which may be plausible given the distracting conditions of a speed dating session. In a laboratory experiment, Jones, Love and Maddox (2006) attempt to disentangle perceptual contrast effects from inferential decision-making that disproportionately weighs recent information. The authors find evidence for both forms of perceptual and inferential recency in visual learning tasks. We turn, however, to a series of extensions to the main analysis that are most easily reconcilable with an explanation involving a contrast effect.

3.4 Extensions

Linearity of the Effect. Most studies of contrast effects in the laboratory examine whether exposure to stimuli with extreme attribute values—exemplars—affect subsequent decisions. This is generally motivated by the belief that the mechanisms underlying contrast effects are activated by extreme representations (e.g., Mussweiler 2003). In this section, we test whether the influence of prior periods on current dating decisions is linear in recent target attractiveness. Due to sample size constraints, we rely on a quintile ranking of attractiveness within groups defined by gender and session.

We first test for the linearity of the influence of previous target attractiveness on current decisions by estimating the following model:

$$Dec_{e,t} = \alpha + \gamma Attract_{e,t} + \Sigma \beta_q D^{q,n}_{e,t-1} + \xi_e + \varepsilon_{e,t} \tag{4}$$

Here, $D^{q,n}_{e,t-1}$ is a dummy variable indicating the attractiveness quintile, q, of the single most recent target ($n = 1$). Again, the model controls for individual-specific variation in decisions, ξ_e, as well as contemporaneous target attractiveness, $Attract_{e,t}$. We report the full sample results in the first column of Table 5, followed by male and female subgroups in the next two columns. The middle quintile is the excluded category.

An initial observation is that the table affirms findings of the earlier analysis. Male evaluators exhibit a contrast effect in their response to unattractive and attractive prior targets. Formally, we can reject the null of equality between the first and fifth quintile coefficients (F=9.59, $p < 0.01$). Female evaluators do not exhibit a contrast effect as evidenced by an inability to reject equality between any pair of quintile coefficients.

Table 5 also indicates that male evaluators respond linearly to prior past targets. A simple measure of linearity, given a partition across quintiles, is to observe whether the magnitude of coefficients across opposing quintiles are equal and of opposite signs. The point estimates roughly suggest linearity in response to recent target attractiveness. More formally, we cannot statistically reject the null that the coefficient for the lowest and highest quintiles is of equal and opposite size (F= 0.16, p= 0.69), or that the coefficient for quintiles 2 and 4 is of equal and opposite size (F= 0.00, p= 0.98).

Influence of Consecutive Exemplar Targets. In the laboratory, subjects are often primed with multiple exemplar images before being assessed for a contrast effect (Kenrick et al. 1989). In principle, it could be that a sequence of exemplar targets, as well as just a single target, could both lead to a contrast effect. It is alternatively possible that the streak of targets operates through a distinct mechanism altogether but is observationally equivalent to a perceptual contrast.

We investigate the influence of consecutive attractive and unattractive targets by first categorizing all twenty-five quintile combinations of recent target pairs. We then estimate a modified version of Equation 4, with dummy variables representing each quintile combination ($n = 2$), excepting the excluded category of consecutive median, or middle quintile, targets. The model is estimated for all daters and then separately by gender.

Table 5

NON-LINEARITY IN EFFECTS OF RECENT TARGET ATTRACTIVENESS ON DATING DECISIONS

	DEPENDENT VARIABLE - DECISION TO DATE (OLS)					
	SINGLE EXEMPLAR Target Population			TWO EXEMPLARS Target Population		
	All (1)	Male (2)	Female (3)	All (4)	Male (5)	Female (6)
Attractiveness	0.107*** (0.010)	0.112*** (0.012)	0.102*** (0.010)	0.108*** (0.008)	0.116*** (0.012)	0.101*** (0.010)
Attractiveness Quintile 1 - Lag (Low Attractiveness)	0.064* (0.025)	0.071** (0.034)	0.056 (0.038)	0.120 (0.073)	0.046 (0.065)	0.161 (0.120)
Attractiveness Quintile 2 - Lag	0.031 (0.028)	0.019 (0.044)	0.044 (0.037)	0.033 (0.078)	-0.042 (0.080)	0.096 (0.106)
Attractiveness Quintile 3 - Lag	X	X	X	X	X	X
Attractiveness Quintile 4 - Lag	0.000 (0.029)	-0.020 (0.041)	0.021 (0.040)	-0.038 (0.084)	-0.138 (0.010)	0.034 (0.117)
Attractiveness Quintile 5 - Lag (High Attractiveness)	0.005 (0.031)	-0.044 (0.043)	0.058 (0.043)	-0.065 (0.103)	-0.222*** (0.081)	0.041 (0.151)
Fixed Effects for Non-Consecutive Quintile Combinations				X	X	X
N	N = 7,200	N = 3,600	N = 3,600	N = 6,728	N = 3,365	N = 3,363
R^2	0.33	0.35	0.31	0.35	0.36	0.33

Notes: Table 5 reports results from tests of the non-linear influence of recent target attractiveness on dating decisions. The dependent variable is a binary variable indicating an evaluator's "yes" / "no" decision to date each round. Target attractiveness is an average rating from two research assistants on a scale from 1-10. Dummy variables indicate the prior (two) target's inclusion in the indicated attractiveness quintile. The median quintile, or median quintile streak, is excluded. Fixed effects control for evaluator specific decisions across all specifications. The first three columns report results of Equation 4 for the sample of pooled, male, and female targets, respectively, while the next three columns report results for the analogous regression for quintile streaks after including fixed effects for all non-consecutive quintile combinations. Regressions are weighted to account for the varying number of targets encountered by each dater. Standard errors are robust and clustered at the target level and are reported parenthetically.

* significant at 10%; ** significant at 5%; *** significant at 1%

This analysis of consecutive exemplar targets, reported in the second set of columns of Table 5, suggests that male evaluators exhibit an even stronger contrast effect after encountering two attractive (-22%, $p < 0.01$) prior targets. Females actually exhibit a large, positive, but imprecisely measured, response to streaks of unattractive prior targets (+16%, not significant), but do not react negatively to highly attractive exemplar streaks. The imprecision of the analysis is due, in large part, to the relative scarcity of quintile streaks.

Amplification of Effect for Moderate Targets. Psychologists find that contrast effects are typically (or most emphatically) triggered when "ambiguous" or moderate stimuli are judged (e.g., Herr, Sherman and Fazio 1983). In this setting, we test for the amplification of contrast effects in the presence of such stimuli by incrementally removing extreme targets from the sample and iteratively re-estimating the effect. Specifically, we compare the influence of recent target attractiveness on current dating decisions using the baseline specification, for the samples purged of targets in the highest and lowest attractiveness deciles.

The effect magnitudes—the size of the estimated coefficients for the attractiveness of the first lagged partner—are reported for each regression in Table 6. For male evaluators the magnitude of the contrast effect is 59% larger once the most attractive and unattractive targets are removed from the sample. This amplification appears driven by the removal of targets in the 70th to 90th and 10th to 30th percentiles. For female evaluators, there is no evidence for a contrast effect for any sub-population of male targets. It is worth noting that the observed effects may also be the mechanical product of a ceiling (or floor) in willingness to date associated with highly attractive (unattractive) targets but not with more moderate targets.

Table 6

EFFECT AMPLIFICATION FOR MODERATE TARGETS

	LAGGED TARGET COEFFICIENT FOR CONTRAST EFFECT REGRESSION		
	Target Population		
	All (1)	Male (2)	Female (3)
Full Sample of Targets	-0.019** (0.008)	-0.035*** (0.011)	-0.004 (0.010)
10th to 90th Percentile of Targets	-0.017* (0.009)	-0.036*** (0.012)	0.006 (0.014)
20th to 80th Percentile of Targets	-0.016 (0.011)	-0.044*** (0.014)	0.012 (0.016)
30th to 70th Percentile of Targets	-0.029 (0.018)	-0.054*** (0.019)	0.004 (0.029)

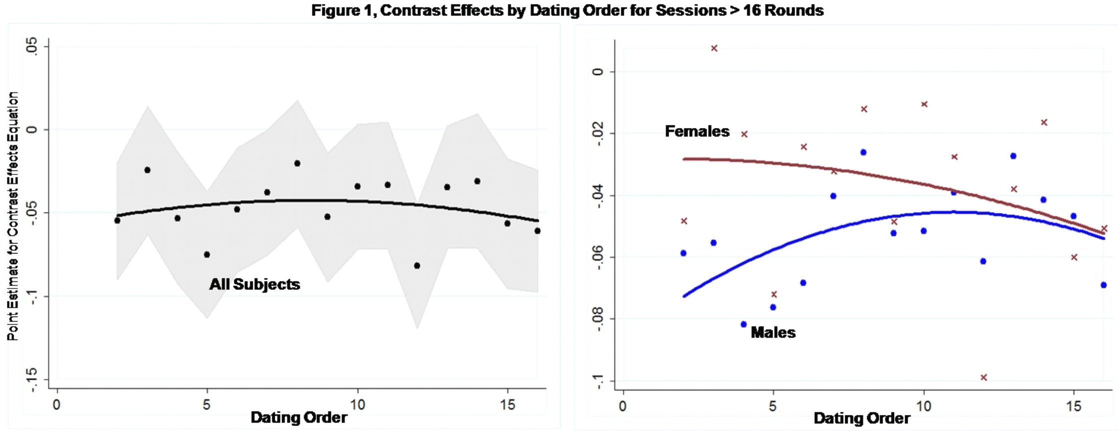
Notes: Table 6 reports results from tests of the amplification of the contrast effect for moderate targets. The magnitude of the contrast effect is indicated by the coefficient for lagged target attractiveness estimated from Equation 2 for varying target populations as described in the text. Regressions are weighted to account for the varying number of targets encountered by each dater. Standard errors are robust and clustered at the target level and are reported parenthetically.

* significant at 10%; ** significant at 5%; *** significant at 1%

Role of Evaluator Experience. Finally, it is natural to ask whether experience moderates the documented non-standard behavior (see Rabin 1998 for discussion). There is mixed evidence in the field for the role played by experience and incentives. Some studies have found that biases, such as the endowment effect, are mitigated by high stakes and experienced agents (e.g., List 2003, List 2004). Others have found that behavioral biases such as the disposition effect among investors (Feng and Seasholes 2005) or loss-aversion among professional golfers (Pope and Schweitzer 2011) are not fully mitigated with experienced agents and high stakes.

In our setting, we can test for the link between experience and the observed contrast effect by examining two measures of decision familiarity. The first is the within-session experience produced by the accumulation of dates. Dating sessions range in length, from 9 to 22 dates and provide varying power for such a test. A second measure of experience is each daters' self-reported dating histories elicited in the pre-session survey. (Of course, because the self-report is a between-dater metric, it is likely to be correlated with a host of other (unobserved) dater attributes.)

Figure 1, Contrast Effects by Dating Order for Sessions > 16 Rounds



We first examine the evolution of contrast effects over the course of each session by estimating the following model, adapted from earlier specifications:

$$Dec_{e,t} = \alpha + \gamma Attract_{e,t} + \Sigma\beta_k(D_k * Attract_{e,t-1}) + \xi_e + \varepsilon_{e,t} \quad (5)$$

Here D_k is a dummy variable indicating the order, k, of each date in a session. The interaction term, $D_k * Attract_{e,t-1}$, is a measure of the order-specific contrast effect. It represents the partial correlation between past attractiveness and current dating decisions by round after controlling for evaluator fixed effects and current target attractiveness. In order to control for compositional effects (i.e., due to variation in session size) we report just the first 16 rounds for sessions of 16 rounds or greater. This reflects a natural demarcation point in the data and it allows us to capture over 75% of the sample. For completeness, in Appendix Figure A1 we display analogous results for all rounds, without controlling for such composition.

Figure 1 illustrate the outcome of this exercise. The left panel displays the contrast effect for all daters across session order with β_k reported on the y-axis and, k, ranging from 2 to 16, on the x-axis. The shaded region represents the 95% confidence interval for each of the estimated coefficients. A quadratic line of best fit is imposed on the scatterplot. For this pooled sample, the contrast effect is relatively stable across rounds. The plot on the right decomposes the effect for male and female daters. The decomposition suggests that the magnitude of the point estimates for males is attenuated by approximately 20% from the first to the second half of the session. Figure A1 displays qualitatively similar results, though estimates for later rounds are subject to imprecision due to small and selected samples.

Table 7

CONTRAST EFFECTS AND SELF-REPORTED EXPERIENCE

	DEPENDENT VARIABLE - DECISION TO DATE (OLS)		
	Target Population		
	All (1)	Male (2)	Female (3)
Attractiveness	0.107*** (0.008)	0.111*** (0.012)	0.103*** (0.010)
Weekly Experience x Attractiveness - Lag	-0.006 (0.014)	-0.008 (0.017)	-0.004 (0.022)
Bi-Weekly Experience x Attractiveness - Lag	-0.021* (0.011)	-0.038** (0.016)	-0.004 (0.014)
Monthly Experience x Attractiveness - Lag	-0.032*** (0.011)	-0.039*** (0.015)	-0.023 (0.018)
Less Than Monthly Experience x Attractiveness - Lag	-0.019* (0.010)	-0.043*** (0.015)	0.000 (0.013)
N	N = 7,200	N = 3,600	N = 3,600
R^2	0.33	0.34	0.31

Notes: Table 7 reports results from tests of whether experience, proxied by self-reported dating history, mediates the influence of recent target attractiveness on dating decisions. The dependent variable is a binary variable indicating an evaluator's "yes" / "no" decision to date each round. Target attractiveness is an average rating from two research assistants on a scale from 1-10. Experience categories indicate self-reported dating frequency from weekly (or sub-weekly) to less than monthly. Fixed effects control for evaluator specific decisions across all specifications. The three columns report results for the sample of pooled, male, and female targets. Regressions are weighted to account for the varying number of targets encountered by each dater. Standard errors are robust and clustered at the target level and are reported parenthetically.

* significant at 10%; ** significant at 5%; *** significant at 1%

We use self-reported measures of dating sophistication from pre-session surveys in a second test of the influence of experience. The survey question specifically asks: "In general, how frequently do you go on dates?" We categorize responses into four, roughly equally proportioned horizons of typical dating frequency—week, bi-weekly, monthly, and less than monthly (i.e., "several times a year", and "almost never"). We estimate the role of experience using the basic specification augmented by the inclusion of interactions between dummy variables indicating experience categories and lagged target attractiveness. Table 7 presents these results. Male daters with weekly dating experience exhibit a contrast effect that is statistically insignificant and is 70 to 80% smaller than those with more limited experience. Consistent with earlier results, females, regardless of sophistication, do not exhibit a contrast effect.

Overall, the analysis of experience suggests at least a partial dampening of contrast effects over the course of accumulated experience. For males, the trajectory of the effect within-session suggests a 20% attenuation by later rounds while correlational analysis of self-reports indicates that those with high dating sophistication are subject to an even more highly attenuated contrast effect.

3.5 Comparison with the Laboratory

A number of studies in the laboratory have examined the role of sequential contrast effects in the assessment of physical attractiveness of strangers (Kenrick and Gutierres 1980; Wedell, Parducci and Geiselman 1987; Kenrick, Gutierres and Goldberg 1989), romantic targets (Weaver, Masland, and Zillmann 1984; Kenrick, Gutierres and Goldberg 1989; Kenrick et al. 1994) and the self (e.g. Cash, Cash and Butters 1983). Kenrick and Gutierres (1980) were the first to investigate contrast effects in the sequential perception of attractiveness and their work constitutes the closest analogue to the present research.[12]

While there are important differences in the specific implementation across settings, the magnitudes of the observed effects of the present analysis of exemplars appear comparable to those found in the laboratory (see Table A2 in the Appendix). We calculate effect sizes in our study by dividing the coefficient estimate for the lagged indicator in the exemplar (streak) regression, reported in Table 5, by the average decision frequency for that gender, and in the laboratory studies, as the percent change between the mean outcome of the treated sample relative to the mean outcome of the untreated sample.

An important divergence between the present study and research in the laboratory is that we seek to label contrast effects as a perceptual *error* by ruling out rational alternative explanations. Research on the perception of social stimuli, such as attractiveness, largely suffers from an identification problem in that one cannot distinguish perceptual errors from

[12] The authors conducted three studies where treatment subjects, primarily undergraduate males, judge the attractiveness of female yearbook photographs after first viewing: (1) a popular television show with attractive female stars, (2) a photograph of a female model in an advertisement, or (3) another yearbook photograph of either a highly attractive or unattractive female.

possible changes in the interpretation of response scales (e.g., Volkmann 1951; Parducci 1963; Biernat, Scherer and Lambert 2009). Consider that for a subject who has just viewed a photograph of an attractive female model, a rating of 4 on a 1-7 scale may connote something different than a rating of 4 elicited from a subject who has just viewed a photograph of an average or unattractive female. Our research makes headway in resolving this issue by using a within rather than between subject design, explicit decision outcomes, and exogenous valuations of each target.[13]

Gender Difference. One notable aspect of our results is that only males exhibit a contrast effect. This asymmetry is consistent with the one laboratory study which investigates female impressions of male attractiveness (Kenrick et al. 1989).[14]

We can speculate as to the causes of this asymmetry. One possibility is that the gender difference is due to procedural details in the administration of the speed-dating sessions. Following established norms (and required by human subjects review), male, but not female, daters rotate from one station to the next after each round. This physical act of approaching a dating partner has been cited as a cause for gender differences in selectivity in speed dating (Finkel and Eastwick 2009). The authors suggest two primary explanations—more positive evaluations by the approaching dater, and heightened self-worth and selectivity of the approached dater—as to why an ostensibly trivial difference in procedure might generate a substantive difference in decision-making. While neither explanation readily explains the gender difference in the tendency to contrast, it is possible that approach norms could contribute in some other manner to the gender difference observed in our setting.

The asymmetry may also be the consequence of gender differences in how attractiveness is assessed. Kenrick et al. (1989) claim that the gender asymmetry associated with contrast effects in the perception of physical attractiveness may be consistent with evolutionary theories of sexual selection (1994). They argue that males and females attend to different aspects of attractiveness— e.g., men attend to "bodily" attractiveness to a greater extent than females. If the evaluation of the research assistants adheres to male, but not female, conceptions, it is possible that our results across gender reflect the particular construction of the objective measure rather than any absence of perceptual contrasts for females.

[13] This is in contrast to early research on the perception of physical stimuli (e.g., Heintz 1950; Sherif and Taub and Hovland 1958; Krantz and Campbell 1961) that finds that subjects, asked to measure the length of a line, the loudness of a sound or the brightness of a color, systematically provide overestimates or underestimates after exposure to extreme lines, sounds, or colors. Given their use of absolute, common-knowledge metrics (e.g. "inches"), this research convincingly differentiates perceptual errors from changes in the interpretation of response scales.

[14] The authors find that exposure to female erotica attenuates male subject ratings of a subsequent photograph as well as subject affection for a romantic partner, but that the same is not true for female subjects exposed to male images (Kenrick et al. 1989).

4 Conclusion

A rich literature in social cognition asserts that perceptions made in sequence are fundamentally relative. In this paper, we examine this claim through an analysis of decisions in the setting of speed dating. After offering evidence that the order of dating targets is conditionally random, we document a negative correlation across dater decisions with respect to perceptions of physical attractiveness for male evaluators. The influence of a prior target's attractiveness on a current dating decision is substantial; for males, it is 31 percent as large as the influence of the current target. We further show that the contrastive influence of past target attractiveness is confined to the most recent target and argue that the differential influence of recent as compared to more distant past targets is evidence for a bias in perception but not consistent with alternative explanations such as learning or the presence of quotas.

We present three additional findings that support our theoretical interpretation and offer insight into the nature of the contrast effect we document. First, we show that the contrast effect is linear in recent attractiveness and may be augmented after encountering multiple attractive or unattractive targets. Second, consistent with laboratory findings, we document an amplification of the effect for current targets of moderate attractiveness. Finally, we demonstrate that experience, both within-session and self-reported, may attenuate, but not fully eliminate, the contrast effect.

While we generally find comparable effect sizes to those found by psychologists in the laboratory, unlike most laboratory studies, we attempt to distinguish between perceptual errors and alternative explanations through the use of a large within-subject sample and by using actual decisions rather than numerical assessments.

It is important to understand whether these results project to other domains involving sequential decisions. Ideally this analysis will be an initial step towards a broader understanding of the role of perceptual biases and sequential context in a broad array of repeated decisions in the field. If contrast effects, of the magnitudes found here and in the laboratory, persist in environments with experienced agents and real stakes, there are implications for decisions ranging from employee hiring and medical diagnoses to policy and investment decisions. Moreover, it is possible that firms or agents might exploit awareness of such a bias in order to shape the decisions of consumers. Future research may elucidate the existence of contrast effects across other domains as well as deepen our theoretical understanding of the factors that shape the size and persistence of such effects.

5 References

Becker, Gary, "A Theory of Marriage: Part I," *The Journal of Political Economy*, Vol. 81, No. 4, pp. 813-846, 1973.

Belot, Michele, and Marco Francesconi, "Can Anyone be 'The' One? Evidence on Mate Selection from Speed Dating," *IZA Discussion Paper*, No. 2377, 2006.

Beggs, Alan, and Kathryn Graddy, "Anchoring Effects: Evidence from Art Auctions," *American Economic Review,* Vol 99, No. 3, pp. 1027-1039, 2009.

Benjamin, Daniel, and Sebastian Brown, and Jesse Shapiro, "Who is 'Behavioral'? Cognitive Ability and Anomalous Preferences," Journal of the European Economic Association, forthcoming.

Bhargava, Saurabh, "Law and Order: Sequential Contrast Effects in Judicial Decisions," *University of Chicago Working Paper,* 2012.

Cash, Thomas, and Diane Cash and Jonathan Butters, "Mirror, Mirror, on the Wall...?," *Personality and Social Psychology Bulletin*, Vol. 9, No. 3, pp. 351-358, 1983.

Choi, Syngjoo, and Shachar Kariv, and Wieland Muller, and Dan Silverman, "Who Is (More) Rational?," *NBER Working Paper*, No. 16791, 2011.

Damisch, Lysann, and Thomas Mussweiler, and Henning Plessner, "Olympic Medals as Fruits of Comparison? Assimilation and Contrast in Sequential Performance Judgments," *Journal of Experimental Psychology: Applied*, Vol. 12, No. 3, pp. 166-178, 2006.

Eastwick, Paul, and Eli Finkel, "Sex Differences in Mate Preferences Revisited: Do People Know What They Initially Desire in a Romantic target," *Journal of Personality and Social Psychology*, Vol. 94, No. 2, pp. 245-264, 2008.

Feng, Lei, and Mark Seasholes, "Do Investor Sophistication and Trading Experience Eliminate Behavioral Biases in Financial Markets?" *Review of Finance*, Vol. 9, pp. 305-351, 2005.

Finkel, Eli, and Paul Eastwick, "Arbitrary Social Norms Influence Sex Differences in Romantic Selectivity," *Psychological Science*, Vol. 20, No. 10, pp. 1290-1295, 2009.

Fisman, Ray, and Sheena Iyengar, and Emir Kamenica, and Itamar Simonson, "Gender Differences in Mate Selection: Evidence from a Speed Dating Experiment," *Quarterly Journal of Economics*, Vol. 121, No. 2, pp. 673-697, 2006.

Fisman, Ray, and Sheena Iyengar, and Emir Kamenica, and Itamar Simonson, "Racial Preferences in Dating," *Review of Economic Studies*, Vol. 75, No. 1, pp. 117-132, 2008.

Fryer, Roland, and Matthew Jackson, "A Categorical Model of Cognition and Biased Decision-Making," *Contributions in Theoretical Economics,* Vol. 8., No. 1, 2008.

Heintz, Roy, "The Effect of Remote Anchoring Points Upon the Judgment of Lifted Weights," *Journal of Experimental Psychology*, Vol. 40, pp. 584–591, 1950.

Higgins, Edward, "Self-Discrepancy: A Theory Relating Self and Affect," *Psychological Review*, Vol. 94, pp. 319–340, 1987.

Hitsch, Guenter and Ali Hortacsu and Dan Ariely, "Matching and Sorting in Online Dating Markets," *American Economic Review*, Vol. 100, No. 1, pp. 130-163, 1987.

Jones, Matt, and Bradley Love, and W. Maddox, "Recency Effects as a Window to Generalization: Separating Decisional and Perceptual Sequential Effects in Category Learning," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 32, No. 3, pp. 316-332, 2006.

Kahneman, Daniel, and Dale Miller, "Norm Theory: Comparing Reality to Its Alternatives," *Psychological Review*, Vol. 93, No. 2, pp. 136–153, 1986.

Kamenica, Emir, "Contextual Inference in Markets: On the Informational Content of Product Lines," *American Economic Review*, Vol. 98, No. 5, pp. 2127-2249, 2008.

Kenrick, Douglas, and Steven Gutierres, "Contrast Effects and Judgments of Physical Attractiveness: When Beauty Becomes a Social Problem," *Journal of Personality and Social Psychology*, Vol. 38, pp. 131–140, 1980.

Kenrick, Douglas, and Steven Gutierres and Laurie Goldberg, "Influence of Popular Erotica on Judgments of Strangers and Mates," *Journal of Experimental Social Psychology*, Vol. 25, No. 2, pp. 159-167, 1989.

Krantz, David, and Donald Campbell, "Separating Perceptual and Linguistic Effects of Context Shifts Upon Absolute Judgments," *Journal of Experimental Psychology*, Vol. 62, No. 1, pp. 35-42, 1961.

Kreps, David, *Notes on the Theory of Choice*, Westview Press, 1988.

Kurzban, Robert, and Jason Weeden, "HurryDate: Mate Preferences in Action," *Evolution and Human Behavior*, Vol. 26, pp. 227-244, 2005.

List, John,."Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics*, Vol. 118, No. 1, pp. 41–71, 2003.

List, John, "Neoclassical Theory versus Prospect Theory: Evidence from the Marketplace," *Econometrica*, Vol. 72, No. 2, pp. 615–625, 2004.

Morse, Stanley, and Kenneth Gergen, "Social Comparison, Self-Consistency, and the Concept of Self," *Journal of Personality and Social Psychology*, Vol. 16, No. 1, pp. 148-156, 1975.

Mussweiler, Thomas, "Comparison Processes in Social Judgment: Mechanisms and Consequences," *Psychological Review*, Vol. 110, pp. 472-489, 2003.

Nickell, Stephen, "Biases in Dynamic Models with Fixed Effects," *Econometrica*, Vol. 49, pp. 1417-1425, 1981.

Parducci, Allen, "Range-Frequency Compromise in Judgment," *Psychological Monographs*, Vol. 77, No. 565, pp. 1-29, 1963.

Pepitone, Albert, and Mark DiNubile, "Contrast Effects in Judgments of Crime Severity and the Punishment of Criminal Violators," *Journal of Personality and Social Psychology*, Vol. 33, No. 4, pp. 448-459, 1976.

Phillips, Phillips, and Donggyu Sul, "Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence," *Journal of Econometrics*, Vol. 137, No. 1, pp. 162-188, 2007.

Pope, Devin, and Maurice Schweitzer, "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes," *American Economic Review*, Vol. 101, pp. 129-157, 2011.

Quattrone, George, and Edward Jones, "The Perception of Variability within In-Groups and Out-groups: Implications for the Law of Small Numbers," *Journal of Personality and Social Psychology*, Vol. 38, pp. 141-152, 1980.

Rabin, Matthew, and Joel Shrag, "First Impressions Matter: A Model of Confirmatory Bias," *Quarterly Journal of Economics*, Vol. 114, No. 1, pp. 37-82, 1999.

Rabin, Matthew, "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics*, Vol. 117, No. 3, pp. 775-816, 2002.

Rabin, Matthew, "Psychology and Economics," *Journal of Economic Literature*, American Economic Association, Vol. 36, No. 1, pp. 11-46, 1998.

Scherer, Laura and Alan Lambert, "Contrast Effects in Priming Paradigms: Implications for Theory and Research on Implicit Attitudes ," *Journal of Personality and Social Psychology*, Vol. 97, No. 3, pp. 383-403, 2009.

Sherif, Muzafer, and Carl Hovland, *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change*, New Haven: Yale University Press, 1961.

Sherif, Muzafer, and Daniel Taub and Carl Hovland, "Assimilation and Contrast Effects of Anchoring Stimuli on Judgments," *Journal of Experimental Psychology*, Vol. 55, No. 2, 1958.

Simonsohn, Uri, and George Loewenstein, "Mistake #37: The Effect of Previously Encountered Prices on Current Housing Demand," *The Economic Journal*, Vol. 116, pp. 175-199, 2006.

Simonsohn, Uri, "New Yorkers Commute More Everywhere: Contrast Effects in the Field," *The Review of Economic and Statistics*, Vol. 88, No. 1, pp. 1-9, 2006.

Simonson, Itamar, and Amos Tversky, "Choice in Context: Tradeoff Contrast and Extremeness Aversion," *Journal of Marketing Research*, Vol. 29, pp. 281-295, 1992.

Stanovich, Keith, and Richard West, "Individual Differences in Rational Thought," *Journal of Experimental Psychology: General*, Vol. 127, No. 2, pp. 161-188, 1998.

Tversky, Amos, and Daniel Kahneman, "Belief in the Law of Small Numbers," *Psychological Bulletin,* Vol. 76, No. 2., pp. 105-110, 1971.

Tversky, Amos, and Daniel Kahneman, "Judgment Under Uncertainty: Heuristics and Biases," *Science,* Vol. 185, pp. 1124-1131, 1974.

Tversky, Amos, and Itamar Simonson, "Context-Dependent Preferences," *Management Science,* Vol. 39, No. 10., pp. 1179-1189, 1993.

Volkmann, John, "Scales of Judgment and Their Implications for Social Psychology," in John Rohrer and Muzafer Sharif (Eds.) *Social Psychology at the Crossroads*, pp. 273-294, New York: Harper, 1951.

Weaver, James, and Jonathan Masland, and Dolf Zillmann, "Effect of Erotica on Young Men's Aesthetic Perception of Their Female Sexual targets,"*Perceptual and Motor Skills,* Vol. 58, pp. 929-930, 1984.

Wedell, Douglas, and Allen Parducci, and R. Edward Geiselman, "A Formal Analysis of Ratings of Physical Attractiveness: Successive Contrast and Simultaneous Assimilation,"*Journal of Experimental Social Psychology,* Vol. 23, pp. 230-249, 1987.

6 Appendix

Table A1

SUMMARY STATISTICS

	Target Population		
	All	Male	Female
Attractiveness (RA)	5.30	5.30	5.31
	(1.48)	(1.42)	(1.55)
Ambitiousness	7.11	7.43	6.79
	(1.77)	(1.77)	(1.72)
Attractiveness	6.48	6.27	6.69
	(1.88)	(1.92)	(1.82)
Fun	6.73	6.56	6.90
	(1.89)	(1.96)	(1.79)
Intelligence	7.78	7.92	7.64
	(1.41)	(1.43)	(1.37)
Shared Similarity	5.54	5.47	5.60
	(2.18)	(2.28)	(2.08)
Sincerity	7.83	7.78	7.88
	(1.58)	(1.69)	(1.45)
Yes Decision	0.42	0.48	0.37
	(0.49)	(0.50)	(0.48)

Notes: Table reports summary stats for attribute ratings and "Yes" decision rates for the pooled sample, as well as by gender separately. Standard deviations are reported parenthetically. Ratings for target attractiveness (RA) are the average ratings of two research assistants on a 1-10 scale, while other attribute ratings are evaluator assessments made in the initial dating round.

Table A2

COMPARISON OF CONTRAST EFFECTS IN THE LABORATORY AND THE FIELD

				CONTRAST EFFECT ESTIMATE (Distortion in Target Judgment Due to Contrast)
TREATMENT	CONTROL	TARGET	JUDGMENT SCALE	RESULT
PERCEPTION OF STRANGERS				
Charlie's Angels (Television Show)	Another show or no TV	Average Female (Yearbook Photograph)	Attractiveness, 1-7	N = 81, -.14 (M) (Kenrick and Gutierres, Study 1, JPSP, 1980)
Farrah Fawcett (Magazine Advertisement)	None	Average Female (Yearbook Photograph)	Attractiveness, Unknown Scale	N = 48, -.25 (M) (Kenrick and Gutierres, Study 2, JPSP, 1980)
Attractive Female Unattractive Female (Yearbook Photograph)	Average Female (Yearbook Photograph)	Average Female (Yearbook Photograph)	Attractiveness, 1-9	N = 98, -.09 (M/F) +.10 (M/F) (Kenrick and Gutierres, Study 3, JPSP, 1980)
Playboy/Penthouse (16 Photographs)	Average Nude Female Abstract Art (16 Photographs)	Average Nude Female (Photograph)	Attractiveness/Desirability, 1-27 (Composite Scale)	N = 196, -.32 (M), -.15 (F) -.19 (M), -.10 (F) (Kenrick et al., Study 1, JESP, 1989)
PERCEPTION OF ROMANTIC PARTNERS				
Attractive Nudes Unattractive Nudes (20 slides + 6 mn Video)	Nature Scenes (20 slides + 6 mn Video)	Romantic Mate	Attractiveness, 0-10 (Multi-category Scale)	N = 46, -.12 (M) +.09 (M) (Weaver et al., PMS, 1984)
Opposite Gender Erotica (Photograph)	Abstract Art	Romantic Mate	Attractiveness/Desirability, 1-27 (Composite Scale)	N = 196, -.16 (M), -.01 (F) (ns) (Kenrick et al., Study 2, JESP, 1989)
PRESENT STUDY				
High Quintile Target Low Quintile Target	Median Dating Target	Dating Target	Yes/No Decision, 0-1	N = 474, n = 7,200, -.09 (M), No Effect (F) N = 474, n = 7,200, +.15 (M), +.15 (F)
High Quintile Target Streak Low Quintile Target Streak	Median Dating Target	Dating Target	Yes/No Decision, 0-1	N = 474, n = 6,728, -.46 (M), No Effect (F) N = 474, n = 6,728, +.10 (M), +.43 (F)

Notes: This table compares laboratory findings with that of the present study. Experimental effect sizes are calculated as the change in the outcome variable in the contrast condition relative to a control condition. The results referring to the present study are estimated coefficients from the regressions of current dating decisions on prior exemplar targets, as well as, exemplar streaks, relative to baseline decision rates (Table 5). N indicates number of evaluators, while n indicates number of interactions. JPSP refers to *Journal of Personality and Social Psychology,* PMS refers to *Perceptual and Motor Skills*, and JESP refers to *Journal of Experimental Social Psychology.*

Figure A1, Contrast Effect by Dating Order for All Sessions