Preliminary. Please do not cite without permission.

Context Dependence in Speed Dating: An Empirical Test of Relative Thinking

Florentin Krämer*
University of Munich
mail@florentinkraemer.com

March 1, 2017

Context influences decision making. However, evidence on the specific underlying mechanism is scarce. Since context dependence in decision making may impact consumers, firms and political agenda setters, exploring the precise mechanism through which choices are distorted is imperative. This paper provides evidence that expanding the utility range of one choice dimension leads decision makers to attach *less* weight to this attribute, in line with the notion of relative thinking (Bushong, Rabin, and Schwartzstein 2015). Context dependence is more pronounced if evaluators are female and survives extensive robustness checks. Implications for policy are discussed.

1 INTRODUCTION

Cognitive psychologists propose that perception depends on context: The same shade of grey is perceived as dark when viewed against a light background and as light when viewed against a dark background. Assuming that context dependence in perception carries over to the domain of preferences implies that choice sets may affect preferences. A number of puzzling phenomena in individual choice can be explained when

*I would like to thank Johannes Abeler, Sandro Ambühl, Thomas Fackler, Dominik Fischer, Felix Klimm, Johannes Maier, Takeshi Murooka, David Schindler, Klaus Schmidt, Lucas Stich, Peter Schwardmann, and Mark Westcott as well as participants of the Micro Workshop at LMU Munich for helpful comments and suggestions. I am very much indebted to Ray Fisman for generously making the speed dating dataset publicly available. I acknowledge funding through the International Doctoral Program *Evidence-Based Economics* of the Elite Network of Bavaria.

allowing for preferences to depend on context, among them attraction effects, preference reversals and certain patterns in intertemporal decision making.

Context dependence essentially means that the composition of the choice set affects the evaluation of its members; hence decision makers may attach different utility values to the same option depending on which other options are in the choice set. Taking such effects into account allows to explain empirical puzzles that cannot be easily reconciled with standard decision theory. One of the most well-known anomalies in this regard are decoy effects, where the presence of an inferior alternative changes choice behavior between two pre-existing options. Models of context-dependent preferences can also shed light on the famous jacket-calculator example advanced by Tversky and Kahneman (1981), consisting of the observation that people are more willing to spend a fixed amount of time in order to save $5 on a $15 rather than on a $125 purchase. In the economic literature, two recent theoretical frameworks endogenize decision weights by relating them to the range of utility levels along a given attribute (Kőszegi and Szeidl 2013; Bushong, Rabin, and Schwartzstein 2015).[1] Though conceptually very similar, these two models disagree about the mapping from utility ranges to decision weights—a feature that makes them prime candidates for testing this particular assumption.

In what follows, I propose a direct empirical test of context dependence (as modeled in Kőszegi and Szeidl (2013) and Bushong, Rabin, and Schwartzstein (2015)), using data from a speed-dating experiment previously reported in Fisman et al. (2006), Fisman et al. (2008) and Bhargava and Fisman (2014).[2] I argue that this experiment provides an ideal setup to test for these types of context effects. First, and most importantly, participants meet a sequence of partners and are asked to evaluate them after

[1] Another class of models is due to Bordalo, Gennaioli, and Shleifer (2012a, 2012b, 2013a, 2013b, 2015). Their basic intuition is that an attribute stands out more the farther it is away from some reference level of that attribute in relative terms. Note that this model differs from Kőszegi and Szeidl (2013) and Bushong, Rabin, and Schwartzstein (2015) in important regards: first, not the utility range along a dimension itself determines how much weight it receives, but rather the comparison of ratios across different dimensions. Second, and related, this implies that different options can have different 'salience rankings', i.e. decision weights might differ across alternatives. For these reasons I focus on the latter class of models and leave empirical tests of Bordalo, Gennaioli and Shleifer's salience theory for future work.

[2] All three papers are based on the same dataset. Fisman et al. (2006) describe gender differences in dating behavior, documenting that men focus more on attractiveness of partners while intelligence is more important for women. In Fisman et al. (2008), the authors examine whether dating behavior exhibits racial biases. They show that same-race preferences are more prevalent for women. In addition, they document determinants of such preferences. Finally, Bhargava and Fisman (2014) explore contrast effects, a phenomenon whereby a potential partner, say, is perceived differently depending on whether or not he or she is preceded by a very attractive man or woman. They show that the bigger the contrast, the lower the likelihood of a positive dating decision for the partner following immediately afterwards. Since contrast effects are a potential confound of context effects, this paper is most closely related to the present work. More details regarding how contrast effects are addressed in the empirical analysis are provided in Section 3.3.

each date, along with deciding whether they would like to meet that person again. The evaluation stage is particularly interesting for my purposes, since it allows to pin down attribute utility values. Second, the sequential nature of speed dating generates within person variation in utility ranges along attributes. Third, the order in which participants speed-date each other is effectively random. This property, in combination with the within-person variation in utility ranges, ensures clean identification of context effects. Finally, dimensions used for ranking potential partners are exogenously given, meaning that they cannot be assumed *ex post*.

The empirical strategy I employ is informed by the central assumptions of focusing (Kőszegi and Szeidl 2013) and relative thinking (Bushong, Rabin, and Schwartzstein 2015) that relate utility ranges to decision weights. At the end of each speed date, subjects report whether they would like to receive the contact details of their respective partners in order to arrange future dates. Due to the simultaneity and privacy of this binary yes/no-decision (which is revealed to the partner only if he/she has indicated 'yes'), it is a dominant strategy to reveal preferences truthfully. The decision to date serves as dependent variable.

In addition to attribute values, I include attribute *ranges* as independent variables in order to understand the effects of context on the likelihood to date. Put more succinctly, this empirical strategy allows for previous observations (partners) to influence the utility of the current partner through their effect on the range of attribute values. Results indicate that decision weights *decrease* in utility ranges, consistent with the framework proposed by Bushong, Rabin, and Schwartzstein (2015). This effect is more pronounced for women than it is for men, and survives extensive robustness checks.

As pointed out before, Kőszegi and Szeidl (2013) and Bushong, Rabin, and Schwartzstein (2015) differ in how they endogenize the decision weights placed on different dimensions. Both models assume that decision weights depend on the range of utility that can be experienced along a given dimension. However, they make the exact opposite assumption on *how* weights are related to the range: Kőszegi and Szeidl (2013) argue that larger ranges attract more attention, and are thus overweighted. In contrast, Bushong, Rabin, and Schwartzstein (2015) assume that fixed differences loom larger when viewed against small ranges, hence decision weights are *decreasing* in the utility range.

To make sense of this discrepancy, it will be helpful to review the main arguments behind these two frameworks. Kőszegi and Szeidl (2013) argue that people focus attention on those dimensions in which they can experience the largest utility gains or losses. All else being similar, it pays to attend to attributes that differ because they generate the largest variability in utility. Roughly speaking, Kőszegi and Szeidl (2013) model behavior as a strategy of "going for large gains and avoiding large losses". This strategy invariably bears the risk of missing fine differences between options. Dertwinkel-Kalt et al. (2016) lend some empirical support to the focusing hypothesis. In a laboratory experiment designed to elicit time preferences, they vary both the distribution of total monetary payments across periods and whether concentrated payments are made earlier or later (there are 9 payment periods in total). Their findings suggest that there is indeed a bias towards concentration in intertemporal choice. Subjects behave relatively more patiently when this results in a stream of small negative consequences followed by a large, concentrated, positive consequence. The converse is also true: when behaving impatiently results in small negative consequences that are preceded by a large positive payoff, subjects will give disproportionate weight to the concentrated advantage and hence behave impatiently, even when compared with a present-biased individual.

Bushong, Rabin, and Schwartzstein (2015) take a different stand: When individuals are asked to compare options along multiple attributes, they will tend to think in relative terms. That is, a small utility difference may appear large when viewed against a relatively small range, and a large utility difference may appear small when evaluated relative to a large range. They argue that several empirical observations are consistent with this notion: the predictions of relative thinking coincide with experimental evidence on attraction effects (Soltani, De Martino, and Camerer 2012), are tentatively supported by evidence on labor supply and wage expectations (Bracha, Gneezy, and Loewenstein 2015), and provide novel insights into "first-of-the-month effects" (Huffman and Barenstein 2005).[3]

In summary, empirical evidence for the specific type of context dependence proposed by these frameworks is mixed. Although intuitively appealing and plausible, Kőszegi and Szeidl (2013) and Bushong, Rabin, and Schwartzstein (2015) fundamentally disagree about the effects of the utility range along one dimension on its decision weight. It is therefore important to shed light on this issue from an empirical perspective. The speed dating experiment allows for a direct test of the central assumption underlying the aforementioned models.

Given that the speed dating setting is relatively specific we might ask whether the findings generalize to other situations. After all, choosing a vacuum cleaner may be an altogether different experience than expressing romantic interest in a potential partner. However, there is no *a priori* reason to believe that individuals should behave differently in other areas of decision making. If anything, context effects may be more pronounced: finding a suitable partner (for a lifetime relationship, or even for a short-

[3] Bushong, Rabin, and Schwartzstein (2015) note that focusing and relative thinking may both shape behavior, despite their apparent contradictions. One way to reconcile these theories would be to assume they influence behavior at different points in the decision process: in order to form a consideration set of a large number of options, individuals may find it useful to eliminate options with many similar attributes and keep only those which they can expect to rank easily. In a second step, they may then use relative thinking to arrive at a final choice. Since I do not observe the entire decision process, but only final choices, the data at hand cannot speak to this hypothesis.

time affair) is a difficult endeavor. It involves a lot of margins that are hard to foresee, yet immensely consequential. If we can be influenced by context in such important choices, it seems probable that context effects are at least as relevant for less involved decisions.

2 THEORETICAL FRAMEWORK

The following section presents a simple model of context-dependent preferences, designed to guide the empirical analysis. It is based on Kőszegi and Szeidl (2013) and Bushong, Rabin, and Schwartzstein (2015).

In the spirit of these two models, I assume that decision makers act on *context-dependent* utility, where the weight attached to each dimension is allowed to depend on other options in the choice set. Formally, let K denote the number of attributes, with a generic attribute indexed by k. Each option $c \in C \subset \mathbb{R}^K$ is a vector of attribute levels: $c = (c_1, \dots, c_K)$.

In line with Kőszegi and Szeidl (2013) and Bushong, Rabin, and Schwartzstein (2015), utility is assumed to be additively separable. Yet instead of maximizing consumption utility $U(c) = \sum_{k=1}^{K} u_k(c_k)$, decision makers act on context-dependent utility $\tilde{U}$:

$$\tilde{U}(c, C) = \sum_{k=1}^{K} g_k u_k(c_k), \tag{1}$$

where $g_k \equiv g(\Delta_k(C))$ denotes the decision weight placed on dimension k and $\Delta_k(C) = \max_{c_k \in C} u_k(c_k) - \min_{c_k \in C} u_k(c_k)$. This reflects the assumption that each g_k is a function of the range of options that are present in the consideration set, as in Kőszegi and Szeidl (2013) and Bushong, Rabin, and Schwartzstein (2015). To account for the sequential nature of the speed dating experiment, I include previous observations in the consideration set (see Assumption 1). That is, all options encountered up to the current partner determine the range of options, and thereby the weight.

Assumption 1 (Consideration Sets). *Consideration sets are time dependent and contain all previously encountered options. Formally,* $C^T = \{c^t\}_{t=1}^{T}$.

Assumption 2 (Decision Weights). *Decision weights depend on the utility ranges of attributes of options in* C^T*:* $g_k^T \equiv g(\Delta_k(C^T))$*, where* $\Delta_k(C^T) = \max_{c_k \in C^T} u_k(c_k) - \min_{c_k \in C^T} u_k(c_k)$.

Assumption 2 extends static context dependence to the sequential speed-dating setup. It states that weights are determined as a function of the difference between the highest and lowest attribute level encountered along a given dimension at a given stage in the sequence. I will refer to this difference as the *range of attribute levels*, or *range* for short. The crucial difference between Kőszegi and Szeidl (2013) and Bushong, Rabin, and Schwartzstein (2015) lies in the way the g_k change as the range increases: assuming that larger ranges attract more focus, as in Kőszegi and Szeidl (2013), implies that $g_k'(\cdot) > 0$. If, however, fixed differences seem bigger the more narrow the range, as in Bushong, Rabin, and Schwartzstein (2015), we would expect $g_k'(\cdot) < 0$. The empirical analysis will concentrate on identifying precisely the effect of larger ranges on the decision weight.

Assumption 3 (Linear Utility)**.** *All attribute utility functions are linear:* $u_k(c_k) = \lambda_k c_k \forall c_k \forall k$.

Linearity is a simplifying assumption that underlies most of the examples and applications in Kőszegi and Szeidl (2013) and Bushong, Rabin, and Schwartzstein (2015). Attribute utility functions are parameterized by λ_k, allowing for utility weights to vary by attribute. Importantly, linearity seems to be reasonable with regard to the data at hand (see Figure 3 and the discussion in Section 4.1).

Assumption 4 (Reference Dependence)**.**

(a) Utility exhibits reference dependence. For a given threshold r_k*, it can be written as*

$$\tilde{U}^T(c, C^T) = \sum_{k=1}^{K} g_k^T m_k(c_k, r_k), \tag{2}$$

where $m_k(\cdot) \equiv u_k(c_k) - u_k(r_k)$.

(b) $\forall c_k \forall k$, *either* $\frac{\partial m_k(\cdot)}{\partial c_k} > 0$ *or* $\frac{\partial m_k(\cdot)}{\partial c_k} < 0$ *or* $\frac{\partial m_k(\cdot)}{\partial c_k} = 0$.

In an attempt to allow for the most general functional form, part (a) of Assumption 4 introduces reference dependence into the setup.[4] The aspiration (or reference) level is denoted by r_k, and may thus vary across attributes. Below, I will argue that a subject's own level of a given attribute determines the reference point, but alternative specifications can be accomodated.[5] Part (b) of Assumption 4 is a monotonicity requirement, ensuring that attribute utility functions are increasing, decreasing, or flat over their entire domain.

Together, Assumptions 3 and 4 imply that

[4] In an unpublished companion paper, Bushong, Rabin, and Schwartzstein (2015) combine their theory of relative thinking with reference dependence in the spirit of Kőszegi and Rabin (2006, 2007).

[5] Subjects rate themselves on all attributes prior to the experiment. Due to self-serving biases, these ratings might be biased upwards. To correct for this potential bias, I use the average rating of all other participants in addition. Empirical support for this approach is given in Section 4.1.

$$m_k(\cdot)\begin{cases} > 0 & \text{if } c_k > r_k, \\ = 0 & \text{if } c_k = r_k, \\ < 0 & \text{if } c_k < r_k. \end{cases} \tag{3}$$

Intuitively, expression 3 asserts that there exist reference attribute values r_k, and receiving exactly the aspired level of a given attribute is neutral in terms of utility. If one's partner ranks above (below) the threshold, this results in positive (negative) utility.

Previewing the estimation strategy, it is useful to consider some comparative static properties. The focus here will be on the effect of a larger range on the likelihood to say yes, which is in turn assumed to be a monotonic and increasing function of utility. To be precise, if for a given attribute level a larger range increases the likelihood to say yes, this supports focusing (à la Kőszegi and Szeidl 2013), while the opposite effect is expected for the Bushong, Rabin, and Schwartzstein (2015) framework to hold true.

Consider again expression (2). Keeping the attribute level constant, the marginal effect of an increase in the range on utility is given by

$$\frac{\partial \tilde{U}^T(\cdot)}{\partial \Delta_k(C^T)} = g_k'(\Delta_k(C^T))m_k(c_k, r_k). \tag{4}$$

Thus, if $m_k(\cdot) < 0$, that is if the evaluator ranks higher on an attribute than her partner, Kőszegi and Szeidl (2013) predict that overall utility decreases since g_k increases in the range. Conversely, overall utility must increase if $m_k(\cdot) > 0$, since the higher weight on dimension k contributes positively to $\tilde{U}$. Due to the opposite assumption on how the range relates to decision weights, Bushong, Rabin, and Schwartzstein (2015) make the exact opposite prediction. To summarize:

Prediction 1 (Impact of Range on Utility)**.**

KS: The likelihood to say yes increases in the range when $m_k(\cdot) > 0$, and decreases when $m_k(\cdot) < 0$:

$$\frac{\partial \tilde{U}}{\partial \Delta_k} = g_k'(\Delta_k(C^T))m_k(c_k, r_k)\begin{cases} > 0 & \text{if } m_k(\cdot) > 0, \\ < 0 & \text{if } m_k(\cdot) < 0 \end{cases} \tag{5}$$

BRS: The likelihood to say yes decreases in the range when $m_k(\cdot) > 0$, and increases when $m_k(\cdot) < 0$:

$$\frac{\partial \tilde{U}}{\partial \Delta_k} = g_k'(\Delta_k(C^T))m_k(c_k, r_k)\begin{cases} < 0 & \text{if } m_k(\cdot) > 0, \\ > 0 & \text{if } m_k(\cdot) < 0. \end{cases} \tag{6}$$

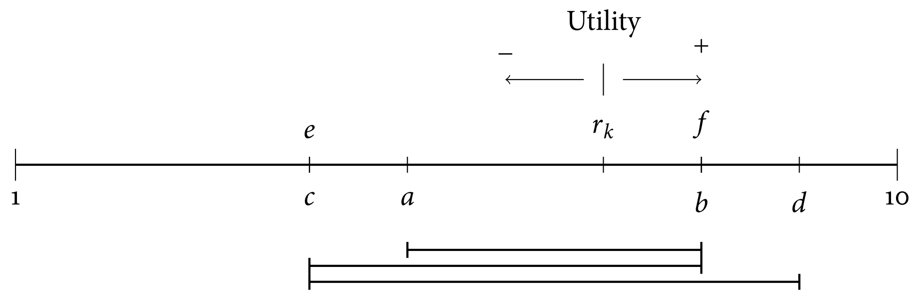


FIGURE 1.—Comparative Statics

Intuitively, the effect of a range increase changes sign depending on whether the attribute's contribution to utility is positive or negative. To illustrate this graphically, suppose that a subject meets a sequence of potential partners (a, b, c, d, e, f). When she meets c, her range is given by the distance between c and b. Continuing in the sequence, her range steadily expands, until she meets partner d. Hence, although e receives the same rating as c, the two decision situations differ in terms of the range (which is $|c-b|$ when deciding about c, but $|c-d|$ when deciding about e). Similarly, b is rated the same as f, but the two situations differ in the range experienced at this point. If ratings between c and e (or b and f) differ depending on the range, holding everything else constant, preferences exhibit context dependence. If the effect of the range differs in sign depending on whether we compare c and e or b and f, preferences exhibit reference dependence. Finally, the sign of the effect in combination with the slope of the utility function determines whether utility is increasing or decreasing in the range.

These comparative static properties inform the estimation strategy described in Section 3.4.

3 EMPIRICAL ANALYSIS

3.1 Institutional Background

Speed dating was invented by Rabbi Yaacov Deyo in 1998 and has spread rapidly since, with only minor variations to the protocol.[6] Its express purpose is to facilitate matching of potential partners by eliminating concerns that a person might not reciprocate one's advances and by enabling people to get to know a large number of other people in a relatively short time span. At a given event, roughly equal numbers of males and females are present. Females are assigned to a table, where they remain seated for

[6] See http://www.nytimes.com/2013/09/29/magazine/who-made-speed-dating.html.

the duration of the event. Men initially pick a table and the pair engages in conversation. After four minutes have elapsed, men rotate to the next table, until every man has talked to every woman. After each speed date, all participants indicate whether they would like to meet the person they have just talked to again. Importantly, the host makes contact information available to the interested parties if and only if both members of a pair agree.

3.2 *Methods and Data*

The data used in this paper were previously reported in Fisman et al. (2006, 2008) as well as Bhargava and Fisman (2014). I will argue below that due to the specific protocol they used, these data are ideally suited to answer questions about salience and focusing. First, however, I will describe the data and the setup.

DATA DESCRIPTION A total of 21 speed dating sessions were conducted at a popular restaurant on the campus of Columbia University, New York City. All environmental aspects, except group size, were held constant. In each session, women were seated at separate tables and remained there during the event. Men were given four minutes to talk to a woman in private, before the organizers indicated the end of the date. Men then rotated to the next table, and this continued until every man had seen every woman. After each speed date, participants were given one minute to rate their counterpart on six attributes and indicate whether they would be interested in receiving their contact details.[7]

All analyses are restricted to 14 out of the 21 sessions. Sessions 6 and 16 have a substantially smaller than average group size (10 resp. 14 participants in total, compared to an average of 28.6 participants). In session 12, participants faced an upper bound on the number of positive decisions to date. In Sessions 18 to 21, participants were encouraged to bring either a book or a magazine, which may have distracted them from the actual speed dating experiment and introduces additional noise.[8]

Participants were also asked to allocate a total of 100 points to attribute categories in order to measure their relative importance. If the total differs from 100 points (presumably due to calculation errors), I renormalize attribute weights.

In sum, this leaves 400 participants (200 male, 200 female) and a total of 6256 observations. Table 1 gives an overview of all sessions that are included in the data analysis.

[7] Of the six attributes used for evaluation, only five will be used for most of the empirical specification. The omitted attribute 'Shared Interests' does not admit evaluation relative to the reference point I consider in Section 3.4, since one's own level of shared interests is not well defined.

[8] Including all sessions in the analysis leaves the main results qualitatively unchanged (see Appendix A.1).

Session ID	Session Date	Male Participants	Female Partipants
1	October 16, 2002	10	10
2	October 23, 2002	16	19
3	November 12, 2002	10	9
4	November 12, 2002	18	18
5	November 20, 2002	10	10
7	March 26, 2003	16	16
8	April 2, 2003	10	10
9	April 2, 2003	20	20
10	September 24, 2003	9	9
11	September 24, 2003	21	21
13	October 8, 2003	9	10
14	October 8, 2003	18	20
15	February 24, 2004	19	18
17	February 24, 2004	14	10

Table 1.—Summary of the Speed Dating Sessions

SETUP The protocol described above closely resembles that of the largest commercial speed dating agency in New York City at the time the experiment was conducted, HurryDate.[9] For the purposes of this paper, it has several attractive features.

First, participants rate each other on six different attributes in short succession. The attributes are attractiveness, sincerity, intelligence, fun, ambition, and shared interests. These ratings can be interpreted as utility values that participants (subjectively) attach to others. Hence, whenever a participant meets another participant who is either better or worse on some attribute than all other participants he has previously dated, his utility range along this attribute increases. The econometric analysis will exploit precisely this within-person variation in utility ranges along attributes.

Second, the order in which participants speed-date each other is effectively exogenous. There may be some scope for choosing whom to date first, but the ensuing (mechanical) rotation ensures that participants do not date in any particular order. This assumption is empirically substantiated by Bhargava and Fisman (2014), who fail to reject the null of random dater order for all six attributes using the same data set.

Third, the data include not only utility (as measured by ratings along the different attributes), but a yes/no decision for each date. The rules specified that upon completing a follow-up survey the day after the experiment, the organizer would distribute email-addresses of other participants if and only if both had indicated that they would

[9] For a more detailed description of the protocol see Fisman et al. (2006).

be interested in further meetings. The presence of a third party should alleviate any concerns regarding strategic behavior on the part of participants, since it is a dominant strategy to reveal one's romantic interest truthfully.

3.3 Preliminary Considerations

SPECIFICATION OF THE ATTRIBUTE RANGE I initially assume that participants have *perfect memory*, that is, they recall every person they have dated during a session. It follows that the range is given by the difference between the best and worst attribute level encountered in the sequence of decisions leading up to the current date. Alternative specifications are possible and serve as robustness checks (see Section 4.4 and Appendix A.1).

For most of the analysis, I leave the range at the first date unspecified. This implies that all regressions will be performed on subsamples starting with the second date. Alternatively, it may be plausible to assume that subjects' own level of a given attribute (c_i) serves as a reference and thus as a starting point for the determination of the range. In this case, the range for the very first date is determined by the difference between c_i and the attribute level of the partner who was encountered first. Results are qualitatively unchanged (see Appendix A.1). Since specifying the attribute range at the first date requires an additional assumption (and introduces one more researcher degree of freedom), I do not pursue this approach.

IMPORTANCE RATINGS Participants in the speed dating events are likely to differ in how much importance they assign to different attributes. Not taking this heterogeneity into account would skew the interpretation of the estimated coefficients below. For instance, a one unit increase in attractiveness may have a larger effect on the likelihood to say yes as a one unit increase in intelligence. Without knowledge of the relative importance of these two dimensions, we could not tell whether effect sizes differ *conditional* on the prior weight put on a given attribute.

Fortunately, the data include importance ratings of the dimensions, assessed prior to the start of the speed dating experience. Participants were asked to distribute points among the six different attributes, with the constraint that the sum of these weights must add up to 100. The results of this exercise are shown in Table 2, separately by gender.

It is apparent that men place a higher weight on attractiveness than women. Conversely, women seem to find ambition more important in potential partners. To account for the heterogeneity in importance ratings, I weight all ratings (and, by implication, all ranges) by the subject-specific rating (standardized to lie between 0 and 1).

	MEN			WOMEN			
	∅	σ	N	∅	σ	N	p-Values
Attractiveness	27.18	14.49	195	17.91	9.67	196	$p < 0.01$
Sincerity	16.21	7.05	195	17.88	6.78	196	$p < 0.05$
Intelligence	19.19	7.03	195	20.89	6.91	196	$p = 0.52$
Fun	17.75	7.11	195	17.54	5.69	196	$p = 0.84$
Ambition	8.65	5.97	195	13.47	5.47	196	$p < 0.01$
Shared Interests	11.01	6.70	195	12.31	5.86	196	$p < 0.10$

Table 2.—Importance Ratings

TIME TRENDS Another concern that may influence the estimated effects comes from the sequential nature of the speed dating experience. As shown in Figure 2, there is indeed a slightly negative time trend in the rate at which participants say yes. This trend is more pronounced for women than it is for men.[10]

Due to the assumption of perfect memory, the attribute range must weakly increase in the number of dates one has encountered. Since this happens concurrently with the overall negative trend in the yes rate, there is a concern that estimating a negative effect of the range on the likelihood to say yes is in fact due to the position of the date, not the range. To eliminate this concern, I include round fixed effects in all subsequent regressions.

RESCALING Due to the sequential setting of the speed-dating experiment, response variables may be subject to rescaling over time. More precisely, this means that context dependence might directly affect stated attribute utility levels instead of influencing only final outcomes (the decision to date). In their work on contrast effects, Bhargava and Fisman (2014) discuss these potential confounds and circumvent the issue by using attractiveness ratings provided by two research assistants, whom they presume not to be influenced by contrast effects.[11]

While an attractive approach in principle, it is unclear why research assistants' ratings should suffer from contrast effects any less than those provided by evaluators. Furthermore, by relying on the average rating of just two research assistants, there is indeed scope for systematic over- or underrating of targets if the order in which targets are evaluated is not perfectly negatively correlated. Their methodology also

[10] In a simple OLS regression, the coefficient on Round is −0.0028 for the whole sample, −0.0027 for men and −0.0029 for women. All three coefficients are statistically significant at the 1%-level.

[11] Bhargava and Fisman (2014) claim that due to using the attractiveness ratings of two research assistants rather than evaluators' own assessments, target values are assessed exogenously (p. 445).

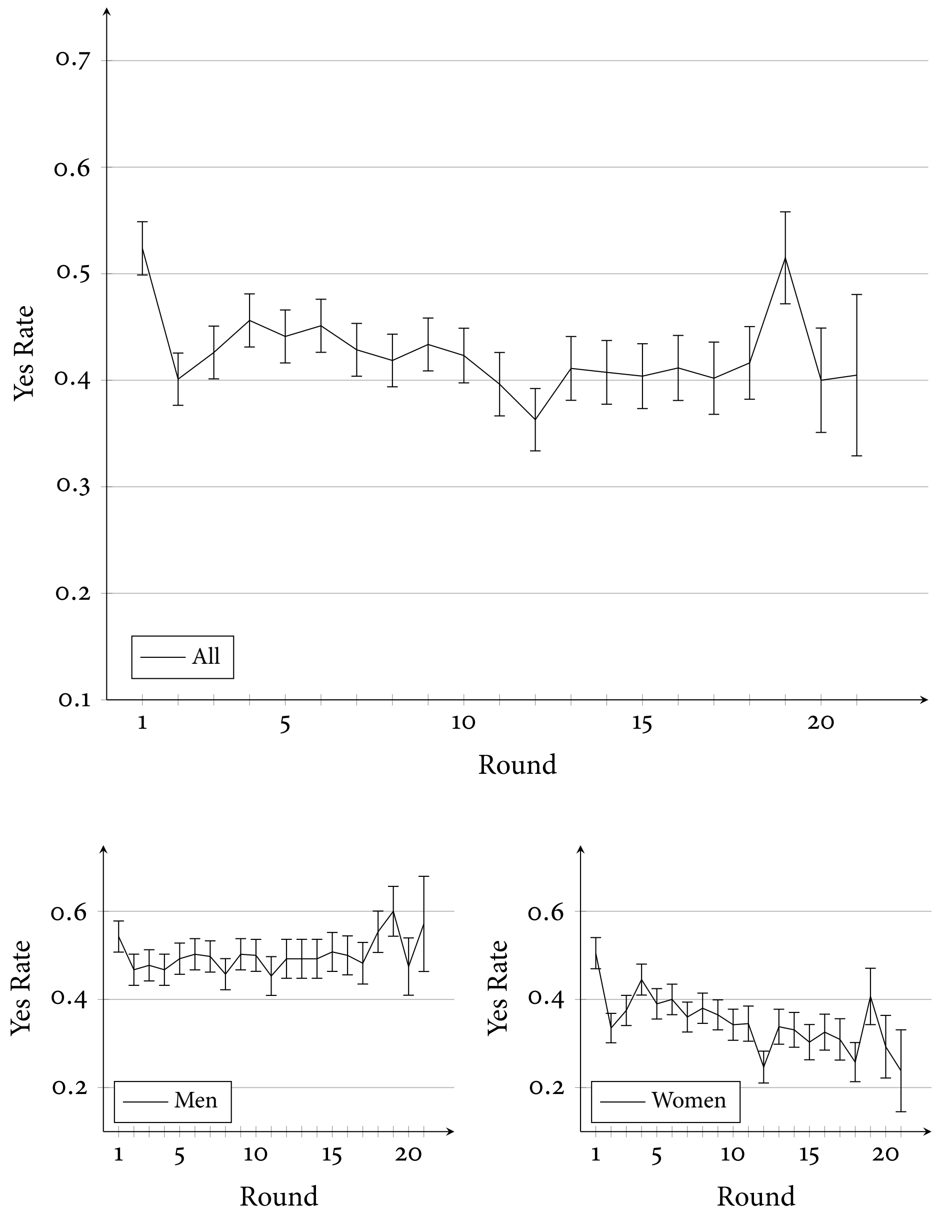


FIGURE 2.—Yes Rates

Notes: Yes rates are averaged over rounds. Error bars display the standard error of the mean, assuming a binomial distribution. Upper panel displays yes rates for all participants. Lower panels are split by gender.

forces them to focus on attractiveness alone (as this is the only attribute that can be assessed without further interaction with the target).

For these reasons I apply an additional, complementary strategy to deal with potential rescaling issues. This strategy rests on the assumption that evaluators cannot be influenced by context effects in the very first round; in which case we can take the assessment of the evaluator who has dated the target in the first round, and apply this rating to assessments of that target in all subsequent rounds. More details regarding this alternative strategy to deal with rescaling and results accounting for this potential confound are reported in Appendix A.1.

CALIBRATION Imagine an evaluator who gives a very high attractiveness rating to the first person he encounters, only to find that as he progresses through the sequence of targets, there are many others whom he deems even more attractive. In this case, his initial evaluation may have been too high *ex post*, and hence persons encountered in earlier periods would be overrated. The opposite could of course also be the case, and I collectively refer to this issue as *imperfect calibration*. If there are systematic patterns of over- or underrating over time, this could compromise the identification of context effects. To deal with this issue, I propose the following method.

First, I assume that for every target j, the 'true' attribute level is given by the average of the ratings of all evaluators except j. I then calculate the deviation of evaluator i's rating to the true attribute level, for each i and all attributes separately. Taking the mean over all deviations by round, I derive a measure for over- or underrating per round.

There is a significant positive trend for seven of these eighteen time series, meaning that subjects in earlier rounds are underrated, while those in later rounds tend to be overrated.[12] However, the presence of such a trend should not bias the results reported below, as any round-specific variation will be picked up by round fixed effects.

CONTRAST EFFECTS To pin down the effect of context on dating decisions, we need to eliminate contrast as a confounding factor. Contrast effects refer to the phenomenon that judgment is comparative: how a judge values the severity of a crime may differ depending on whether she has been exposed to a particularly terrifying crime beforehand, or whether the case is preceded by a similar crime (Pepitone and DiNubile 1976).[13] Considering the speed dating experiment, a person's attractiveness may be judged less favorably when preceded by a very attractive person. However,

[12] There are six attributes (attractiveness, sincerity, intelligence, fun, ambition and shared interests) and three subsamples (all, men, women) for a total of eighteen time series.

[13] For a summary of contrast effects, see Bhargava and Fisman (2014) and the references therein.

an attractive person is also very likely to expand the range, and therefore we cannot discriminate whether the less favorable rating is due to contrast or to the larger range.

Contrast effects have been studied by Bhargava and Fisman (2014), using the same dataset. They document that the more attractive the previous partner was, the less likely it is that the evaluator decides to date the current partner. They also show that such contrast effects are transient, i.e., they are driven to a large extent by the partner immediately preceding the current one. In order to eliminate this confound, I follow their methodology by including lagged attribute values in all regressions. The interpretation of the results will thus be conditional on the presence of potential contrast effects. Put differently, any effect of the attribute range on the decision to date should be regarded as an effect *in addition* to contrast effects.

3.4 Empirical Strategy

To identify the effect of the attribute range on the likelihood to say yes, it is important to distinguish whether the absolute attribute level is regarded as a 'good' or as a 'bad'. This is because the attribute range effectively influences the decision weight placed on a given dimension, and hence we should expect the effect to change sign depending on whether the absolute level is above or below some threshold.

Put more succinctly, if subject i regards partner j as attractive, an increase in the range along that dimension will make it either more or less likely that i says yes to j, depending on whether the decision weight placed on this dimension increases or decreases. Conversely, when j is regarded as unattractive, changing the decision weight must have precisely the opposite effect on the final decision.

As a largely realistic first pass, I consider the evaluator's own level along an attribute as a natural threshold (Figures 3 and 4, described in more detail below, lend support to this approach). Partners rated above one's own level are hence regarded as desirable, whereas a rating below one's own decreases desirability.

After each date, participants are asked to rate their counterparts on six attributes, using a Likert scale from one to ten. Besides these attribute ratings, the data contain ratings of the relative importance of dimensions. These were assessed prior to the speed dating sessions. I use importance-weighted attributes and ranges in the subsequent regressions to account for heterogeneity across participants' utilities derived from a change in a given dimension.

This leads to a model of the form

$$
\begin{aligned}
\text{Decision}_{ijt} = \alpha_i + \gamma_t + \sum_{k \in K} \beta_{k1} \cdot \text{Range}_{ikt} \\
+ \sum_{k \in K} \beta_{k2} \cdot \mathbb{I}(\text{Rating}_{ijkt} > \text{Threshold}_{ik}) \cdot \text{Range}_{ikt} \\
+ \sum_{k \in K^+} \beta_{k3} \cdot \text{Rating}_{ijkt} + \sum_{k \in K} \beta_{k4} \cdot \text{Rating}_{ijk,t-1} + \varepsilon_{ij},
\end{aligned} \tag{7}
$$

where $\mathbb{I}$ is an indicator function assuming the value one if the condition in parentheses evaluates to 'true' and zero otherwise, K includes the subset of attributes for which context effects can be identified (in a sense made precise in Section 4.1), K^+ = {Attractiveness, Sincerity, Intelligence, Fun, Ambition, Shared Interests}, α_i contains subject-specific fixed effects, γ_t are round fixed effects, and $\text{Rating}_{ijk,t-1}$ captures lagged attribute values in order to account for contrast effects. I estimate this model using a linear probability specification with standard errors clustered on the partner level.

All observations are weighted by the inverse of the number of partners encountered in the course of a given session. This ensures that subjects in larger sessions do not receive undue weight.

4 RESULTS

4.1 *Attribute Utility Functions & Thresholds*

The estimation strategy relies on the assumption that there exist reference attribute levels separating utility into 'good' and 'bad'. Context effects of the type proposed by Kőszegi and Szeidl (2013) and Bushong, Rabin, and Schwartzstein (2015) are identified if the effects of the utility range on the decision to date differ in sign depending on whether the evaluated person is rated above or below the reference level.

To get a better sense of the utility functions for each attribute, Figures 3 and 4 present results of five separate regressions of the decision to date on each attribute, including time and individual fixed effects. The dashed line plots linear predictions for every possible level of the attribute, and corresponding 95% confidence intervals are shown as grey areas.

Horizontal red lines depict *average* yes rates. I assume that an attribute is perceived as a good if it leads to an above average propensity to date, and as a bad if it entails a below average willingness to say yes to a date. Given this assumption, we can ask which attribute level marks the threshold that separates utility. For future reference, I show two candidate thresholds, drawn as vertical lines: subjects' own rating of an attribute, averaged over all subjects (green dashed line); and the average consensus rating of subjects, explained in more detail below (blue dashed line).[14]

Figures 3 and 4 document two key insights: first, sincerity, intelligence and ambition seem to have very little predictive power for the decision to date. In fact, the predicted effect is almost flat, implying that the levels of these attributes matter little for dating decisions. Second, the propensity to date increases in attractiveness and fun. The relationship is approximately monotonic. Importantly, the reference level is very well approximated by the average consensus rating, which lends support to the estimation strategy discussed in Section 3.4.[15] These observations are summarized in the following result:

Result 1 (Attribute Utility Functions & Thresholds)**.**

(a) Sincerity, intelligence and ambition possess little to no predictive power for the decision to date.

(b) Utility increases monotonically in attractiveness and fun almost everywhere. Attribute levels above the average consensus rating are perceived as 'good', while those below are perceived as 'bad'.

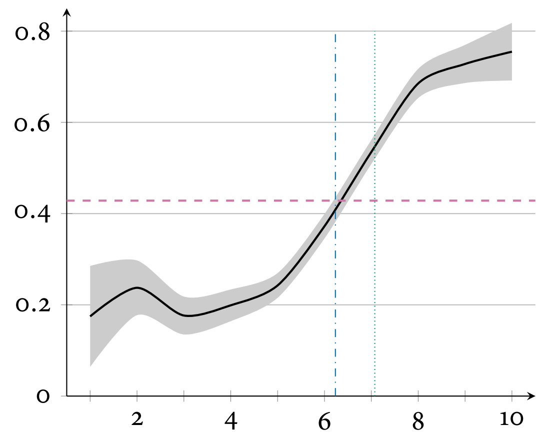
Result 1 states that the predicted effect on the decision to date is constant over attribute levels for three dimensions (sincerity, intelligence and fun), while it is increasing in levels of attractiveness and fun. By implication, the identification strategy proposed in Section 3.4 will only be able to pick up differences in the signs of range effects for the latter attributes. All subsequent analyses will therefore focus on attractiveness and fun.

4.2 Context Dependence

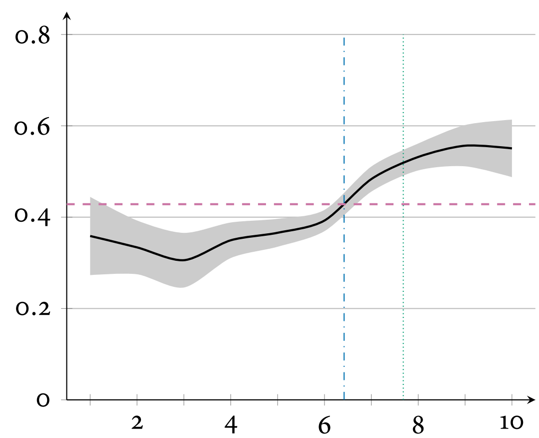
Tables 3 and 4 present the main result. The dependent variable is the decision to date. All regressions include round fixed effects to account for the slight negative time trend in the dependent variable. In addition, they control for absolute attribute levels and lagged absolute attribute levels in order to identify possible contrast effects. Subject-specific fixed effects are included to capture any remaining heterogeneity among participants.

[14] Own attribute ratings were elicited via the question: "How do you think you measure up? Please rate your opinion of your own attributes, on a scale of 1–10 (be honest!)", followed by a list of five attributes (excluding 'Shared Interests').

[15] Given that ratings of one's own attribute levels are likely subject to self-serving biases, it is unsurprising that the blue and green lines do not overlap. A somewhat more surprising finding is that even though individuals consistently report exaggerated own ratings, their judgment seems to be accurately guided by the perceptions of others. However, a more detailed analysis of this phenomenon is beyond the scope of this paper and should therefore be regarded as speculative.



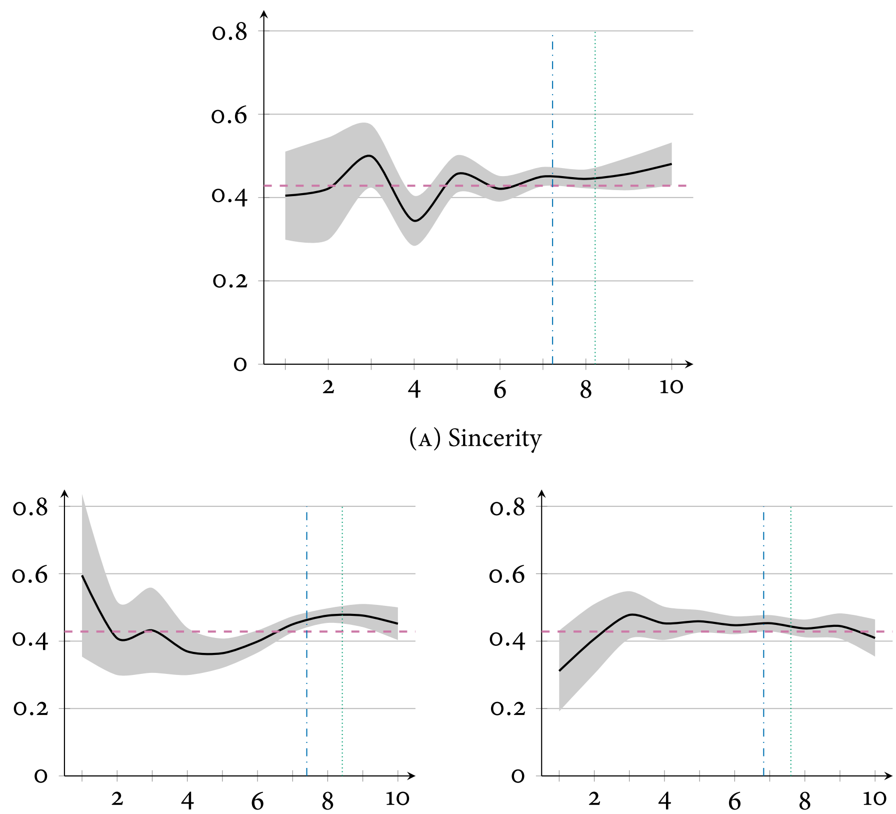
(A) Attractiveness



(B) Fun

FIGURE 3.—Marginal Effects of Attribute Levels on Dating Decision (Attractiveness & Fun)

Notes: Vertical axes display average predicted values of the dependent variable (decision to date) for different attribute values. Plotted values are derived from regressing the decision to date on all five attributes, including time and individual fixed effects. Marginal effects are obtained post estimation for each attribute separately. Regressions are weighted by the number of observations in each session. Standard errors are clustered on the partner level. Grey areas are 95% confidence intervals. Horizontal red lines depict average yes rates, vertical blue lines (dash dotted) are average consensus ratings of the respective attribute, and vertical green lines (dotted) are average own ratings of the attribute.



(A) Sincerity

(B) Intelligence

(C) Ambition

FIGURE 4.—Marginal Effects of Attribute Levels on Dating Decision (Sincerity, Intelligence, & Ambition)

Notes: The same notes as for Figure 3 apply.

All specifications rely on the ratings made by participants at the end of each date. In Table 3, one's own attribute rating is used as threshold for the definition of the indicator variable. That is, the indicator is set to one whenever the partner's attribute level weakly exceeds one's own level of that attribute (assessed prior to the start of the speed dating sequence), and to zero otherwise.

Table 4 differs from Table 3 only in that it uses the consensus rating of all other evaluators as threshold for the indicator variable. The same methodology has been employed by Fisman et al. (2006) to account for potential self-serving biases in the assessment of one's own attribute level.

Since consensus ratings appear to be cleaner both in terms of methodology and by way of indicating the correct reference level (see Figures 3 and 4), the following analysis will focus on the results in Table 4.

Recall that focusing in the spirit of Kőszegi and Szeidl (2013) predicts that larger ranges attract more attention, and therefore the effect of the range on the decision to date should be positive given that utility of the attribute is above the reference level. Conversely, the effect will be negative if below the threshold. This is not what we see: in fact, for attribute levels below the reference level, the effect is *positive* (for both attractiveness and fun). The range effect for attribute levels above the reference level is given by the sum of the coefficients on the attribute range and the interaction between the attribute range and the indicator for being above the reference level. For both attractiveness and fun, this effect is negative. However, the effects are jointly insignificant and thus indistinguishable from zero.[16]

To probe further into these results, columns (2) and (3) split the sample by gender.[17] There are pronounced gender differences: below the reference level, both men and women are more likely to say yes the larger the range. However, only female participants react to range increases above the reference level as predicted by Bushong, Rabin, and Schwartzstein (2015). The joint effect is significantly negative in this case ($F(1, 199) = 6.55$, $p < 0.05$). It is insignificant for men.[18]

[16] Results are qualitatively unchanged when controlling for a linear time trend instead. The joint effects on attractiveness in columns (1) and (3) of Table 3 and column (3) of Table 4 are slightly less precisely estimated: conditional on the partner being above the reference level, an increase in the attractiveness range is significantly negative on the 10% level in specification (1) of Table 3, insignificant in specification (3) of Table 3, and significantly negative at the 5% level in specification (3) of Table 4. Magnitudes of coefficients are very similar.

[17] Psychologists and economists have identified several behavioral traits along which men and women differ. The evidence is not always conclusive: in the domains of risk aversion and altruism, for instance, there does not appear to be a robust and replicable gender gap, while the differences are rather large with respect to attitudes towards competition (for an excellent survey, see Niederle 2016). To the best of my knowledge, this paper is the first to find a gender gap in context dependence, speaking to the domain-specificity of such differences.

[18] Note that these results are not distorted by differential utility functions of male and female participants. Figure 6 in Appendix A.2 shows that both men and women exhibit utility functions that are

	Own Rating		
	(1)	(2)	(3)
	All	Male	Female
Attractiveness Range	0.0696***	0.0955***	0.0264
	(0.0236)	(0.0266)	(0.0481)
More Attractive	0.281***	0.296***	0.278***
	(0.0240)	(0.0340)	(0.0352)
More Attr. × Attr. Range	-0.120***	-0.133***	-0.112***
	(0.0219)	(0.0270)	(0.0399)
Fun Range	0.00982	-0.0272	0.0408
	(0.0348)	(0.0520)	(0.0483)
More Fun	0.0947***	0.110***	0.0708*
	(0.0253)	(0.0342)	(0.0368)
More Fun × Fun Range	-0.0111	-0.0284	0.0120
	(0.0290)	(0.0405)	(0.0422)
Constant	-0.757***	-0.810***	-0.725***
	(0.0608)	(0.0819)	(0.0865)
Attribute Level Controls	Yes	Yes	Yes
Contrast Effect Controls	Yes	Yes	Yes
Round Fixed Effects	Yes	Yes	Yes
Observations	4710	2401	2309
R^2	0.552	0.554	0.539
Adjusted R^2	0.509	0.508	0.490

Table 3.—Context Dependence (Own Rating)

Notes: Linear probability model with standard errors clustered on the partner level (in parentheses). In all regressions, the dependent variable is the Decision to Date. All observations are weighted by the inverse of the number of observations per subject; ratings are adjusted using importance weights. Attribute ratings are on a scale from 1 to 10. More Attractive and More Fun are indicator variables set to one if the partner surpasses the evaluator's own level of the respective attribute, as judged by the evaluator him- or herself.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	Consensus Rating		
	(1) All	(2) Male	(3) Female
Attractiveness Range	0.119***	0.116***	0.112***
	(0.0287)	(0.0350)	(0.0415)
More Attractive	0.125***	0.0963***	0.176***
	(0.0236)	(0.0369)	(0.0323)
More Attr. × Attr. Range	-0.127***	-0.0900***	-0.226***
	(0.0210)	(0.0269)	(0.0357)
Fun Range	0.0641*	0.0279	0.100**
	(0.0373)	(0.0551)	(0.0491)
More Fun	0.0627**	0.110***	-0.00420
	(0.0263)	(0.0370)	(0.0355)
More Fun × Fun Range	-0.0951***	-0.112***	-0.0556
	(0.0278)	(0.0392)	(0.0390)
Constant	-0.967***	-0.967***	-0.975***
	(0.0602)	(0.0838)	(0.0840)
Attribute Level Controls	Yes	Yes	Yes
Contrast Effect Controls	Yes	Yes	Yes
Round Fixed Effects	Yes	Yes	Yes
Observations	4710	2401	2309
R^2	0.538	0.539	0.531
Adjusted R^2	0.494	0.492	0.482

TABLE 4.—Context Dependence (Consensus Rating)

Notes: Linear probability model with standard errors clustered on the partner level (in parentheses). In all regressions, the dependent variable is the Decision to Date. All observations are weighted by the inverse of the number of observations per subject; ratings are adjusted using importance weights. Attribute ratings are on a scale from 1 to 10. More Attractive and More Fun are indicator variables set to one if the partner surpasses the evaluator's own level of the respective attribute, as judged by all other subjects.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

These effects are large and economically significant. A one point increase in the utility range increases the likelihood to say yes by 11 to 12 percentage points if the target is *less* attractive than the evaluator. Conversely, a one point range increase decreases the likelihood to say yes by 12 percentage points if the target is *more* attractive than the evaluator and the evaluator is female.

Results for fun exhibit overall similar patterns. However, they are less precisely estimated (all interaction effects are jointly insignificant). A possible explanation lies in the considerably flatter slope of the utility function relative to attractiveness (see Figure 3). In combination with the smaller effect size, this makes it more difficult to pick up the effect.

Result 2 (Context Dependence)**.**

(a) The decision to date increases in the indicators for 'more attractive than self' and 'more fun than self', which implies that attribute levels in excess of one's own are valued positively.

(b) Results for both attractiveness and fun are qualitatively consistent with Bushong, Rabin, and Schwartzstein (2015), whereby fixed differences loom smaller the larger the range.

(c) There are marked gender differences: female participants are more susceptible to context effects than male participants.

Part (a) of Result 2 supplements and substantiates the graphical analysis presented in Figure 3. Taken together, these observations imply that utility increases over the relevant range of attribute levels.

Part (b) states that in the specific context considered here, Bushong, Rabin, and Schwartzstein (2015) has more traction than Kőszegi and Szeidl (2013). When deciding about whom to date, individuals appear to view differences in attributes across potential partners relative to the range they have experienced up to this point. Since a given utility difference appears smaller when evaluated relative to a large range, decision weights decrease in utility ranges. This results in the finding that a larger range increases the propensity to date when the evaluated partner ranks below some reference level along this attribute, and *vice versa* if ranked above. Yet this is not to say that relative thinking is the *only* determinant of behavior in this context. Rather, future research should identify whether there exist conditions that lead individuals to selectively focus on a subset of dimensions, using the heuristic by which larger ranges imply higher potential utility gains. If this were true, it is possible that both focusing

monotonically increasing in attractiveness almost everywhere. In addition, even though the average yes rate is significantly different, the consensus rating approximates the reference level very well.

and relative thinking are decisive determinants that operate at different stages of the decision process.

Finally, part (c) draws attention to the fact that there are marked gender differences in behavior according to the regressions using the consensus rating of one's own attribute level. Since the effect is stronger for men when using the 'own level'-specification (yet still jointly insignificant), this could be an artifact of incorrectly specifying the reference level. However, it could also point to a true gender difference that remains to be explained.

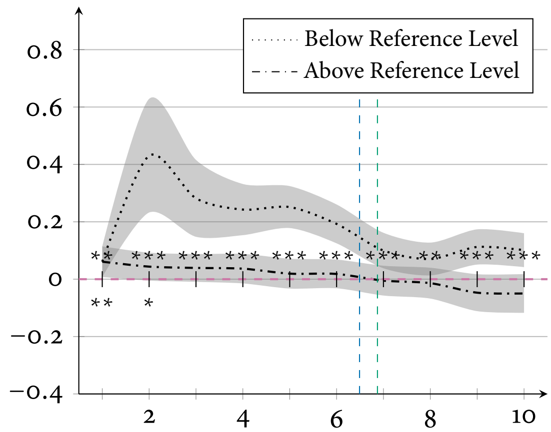
4.3 Uniqueness of the Reference Level

As pointed out before, the estimation strategy rests on the assumption that there exists a reference level of utility. To give substance to this assumption, Figures 3 and 4 show that indeed there appears to be a switch at the intersection of the average yes rate, which can be interpreted as the utility/disutility-threshold, and the average consensus rating. The fact that the estimated utility function passes through the intersection point and is monotonic suggests that the consensus rating serves as a natural reference level for relative utility judgments.

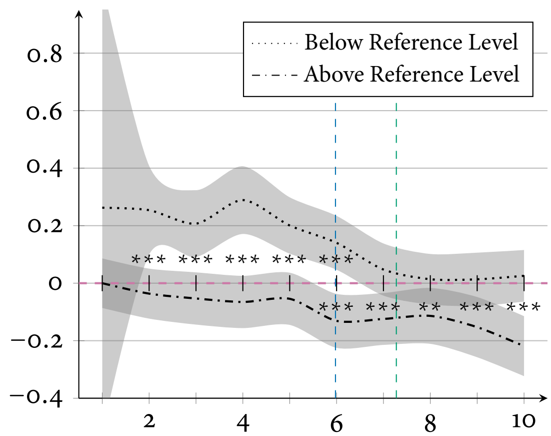
In order to show that this reference level is unique, I conduct placebo tests. The basic idea is as follows: if (i) there exists a utility reference level, and attribute values below this reference level induce disutility, while values above provide utility, and if (ii) the effect of the attribute range on utility differs according to whether the evaluated person is above or below the reference level, there must be a reference level at which both effects are simultaneously discernible. If this is not the case, the location of the reference level should not matter. Hence we would expect to observe similar effect sizes for randomly chosen placebo reference levels.

Figure 5 plots the results for attractiveness. I run regressions of the same type as in Section 4.2, except for one crucial difference: I hold the interaction reference level for fun constant at the consensus rating, while varying the reference level for attractiveness from 1 to 10. The dotted line then plots the coefficients on the attribute range for all placebo reference level values, conditional on the attribute level being *below* this reference level. Conversely, the dash-dotted line plots coefficients on the attribute range conditional on being *above* the reference level. The grey areas depict 95% confidence intervals.

In addition to the coefficients, I show the values of the average consensus rating and the average own rating of the specific attribute (blue and green vertical lines). Asterisks above the null-effect line (marked in red) indicate whether coefficients of the below reference level attribute range are significantly different from zero; analogously for asterisks below the red line and coefficients of the above reference level attribute range.



(A) Men



(B) Women

FIGURE 5.—Placebo Tests of Attractiveness Ranges

Notes: Figure displays regression coefficients of the decision to date on attribute range (dotted line) and the sum of attribute range and the interaction between attribute range and a dummy (dash-dotted line). The dummy is set to one if the attribute level is higher than $k = 1, 2, \ldots, 10$ and set to zero otherwise, such that each line consists of coefficients from 10 separate regressions. Grey areas are 95% confidence intervals. The horizontal axis tracks the value of the threshold k. The two vertical lines indicate the averages of the consensus rating of the attribute (blue line) respectively subject's own rating of the attribute (green line). Stars above the red line indicate whether coefficients of the dotted line are significantly different from zero; analogously for stars below the red line and the dash-dotted line. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To interpret the findings, it is useful to recall the theoretical predictions. Presuming that subjects evaluate partners according to how well they perform relative to some reference level, Bushong, Rabin, and Schwartzstein (2015) predict that the likelihood to say yes will decrease when the utility range along a dimension increases if and only if the partner's utility level on that attribute is above the reference level. Conversely, because an increase in the range leads subjects to focus less on small utility differences, an attribute's disutility will be underweighted when it is below the reference level. Hence the prediction that a range increase will lead to a higher propensity to say yes when the evaluated subject is below the reference level, and to a lower propensity when it is above. If there were indeed a unique threshold level of utility, we would therefore expect to find a significant *below* reference level range coefficient for values up to some reference level, a significant *above* reference level coefficient for values above, and a null effect for all other levels of these two coefficients.

Due to the marked gender differences reported in the main part, performing placebo tests separately by gender is informative. Panel 5b provides evidence that for female evaluators, there is indeed a unique reference level of attractiveness that separates the utility scale into areas of utility and disutility. It is apparent that exactly at the consensus rating (marked in blue), there is a switch both in the magnitude and the significance of the coefficients on the range, conditional on being above or below the reference level. For reference levels up to 6, the *below* reference level range coefficient is highly significant and positive, whereas for reference levels 6 up to 10, the *above* reference level coefficient is highly significant and negative.[19]

Result 3 (Uniqueness of the Reference Level)**.**

(a) In case the evaluator is female, there exists a unique reference level of attractiveness which coincides with the average consensus rating.

(b) For male evaluators, the uniqueness property does not hold: many reference levels are consistent with the proposition that the effect of the attribute range differs depending on whether the evaluated partner ranks below or above the reference level.

(c) Results for the attribute fun *are qualitatively consistent with a unique reference level, but the coefficients are less precisely estimated (cf. Appendix A.3).*

[19] Performing the same type of analysis on the combined sample, Panel 7a in Appendix A.3 shows a pattern that is qualitatively consistent with this expectation, but fails to hold up quantitatively. In particular, there is no apparent change in both the magnitude and the significance of the coefficients at some value between the consensus rating and the own rating. Furthermore, the below reference level coefficient remains significantly positive even for values above the reference level. Recalling that there appears to be no generalized context dependence across genders (cf. Table 4), this finding is expected.

To put this in plain words, Panel 5b tells us that if an econometrician were to set an arbitrary reference level of the attribute attractiveness, she would find that in 8 out of 10 cases, increasing the range influences the outcome variable either positively or negatively, but not both.[20] Only in one case, namely when the reference level is set at 6, will an increase in the range have both a positive effect when the evaluated attribute is below the reference level and a negative effect when it is above, and so this must be the unique reference level separating disutility from utility.[21]

The intuition is fairly straightforward: if an evaluated partner is above the placebo reference level at arbitrary levels below the true reference level, range increases should have no impact on the decision to date, because effects for attractive and unattractive evaluated persons will cancel each other. Hence we should not expect to see an above reference level range effect. Conversely, individuals rated below the placebo threshold are unambiguously unattractive, and so we expect to see an unambiguous range effect. Naturally, the same mechanism applies to the area to the right of the 'true' reference level.

4.4 Consideration Set Composition

Recall that by Assumption 1, all previous dates are included in the consideration set at the time individuals decide about their current partner. The attribute range is hence determined by taking the largest difference between attribute values encountered from the first date to the current one. This is paramount to assuming that context effects are not transient, but permanent. However, it requires that subjects have perfect memory, which may be unnecessarily demanding.

Table 5 reports regressions that vary in memory length. As reference case, column (1) assumes perfect memory. Columns (2) to (4) successively implement shorter memory lengths, such that in specification (4) the range is determined by taking only the previous partner as a reference.

Table 5 indicates that context dependence is robust to varying the composition of the consideration set. Qualitatively, results are unchanged. Effect sizes differ across specifications, but not substantially. For a memory length of 3 respectively 2 partners, (columns (2) and (3)), the effect of attractiveness range conditional on the partner being more attractive is significantly negative ($F(1, 399) = 3.02$, $p < 0.1$; $F(1, 399) = 2.72$, $p < 0.1$).

[20] At a reference level of 1, the effect cannot be estimated reliably for lack of power.

[21] Note that due to the coarseness of the evaluations, which were constrained to whole numbers, it is not meaningful to use finer divisions of the reference level. This implies that estimates for the range coefficients are lumpy, and the lines connecting these estimates are merely interpolations between adjacent points.

	Consensus Rating			
	(1) Perf. Mem.	(2) 3 Partners	(3) 2 Partners	(4) 1 Partner
Attractiveness Range	0.119***	0.0787***	0.0853***	0.124***
	(0.0287)	(0.0254)	(0.0245)	(0.0307)
More Attractive	0.125***	0.0930***	0.0836***	0.0751***
	(0.0236)	(0.0216)	(0.0217)	(0.0201)
More Attr. × Attr. Range	-0.127***	-0.119***	-0.124***	-0.165***
	(0.0210)	(0.0241)	(0.0284)	(0.0421)
Fun Range	0.0641*	0.131***	0.0993***	0.0980***
	(0.0373)	(0.0297)	(0.0276)	(0.0308)
More Fun	0.0627**	0.0594**	0.0377*	0.0332
	(0.0263)	(0.0240)	(0.0228)	(0.0213)
More Fun × Fun Range	-0.0951***	-0.131***	-0.103***	-0.142***
	(0.0278)	(0.0327)	(0.0359)	(0.0494)
Constant	-0.967***	-0.944***	-0.905***	-0.865***
	(0.0602)	(0.0597)	(0.0582)	(0.0555)
Attribute Level Controls	Yes	Yes	Yes	Yes
Contrast Effect Controls	Yes	Yes	Yes	Yes
Order Fixed Effects	Yes	Yes	Yes	Yes
Observations	4710	4710	4710	4710
R^2	0.538	0.537	0.536	0.538
Adjusted R^2	0.494	0.493	0.492	0.494

Table 5.—Consideration Set Composition

Notes: Linear probability model with standard errors clustered on the partner level (in parentheses). In all regressions, the dependent variable is the Decision to Date. All observations are weighted by the inverse of the number of observations per subject; ratings are adjusted using importance weights. Attribute ratings are on a scale from 1 to 10. More Attractive and More Fun are indicator variables set to one if the partner surpasses the evaluator's own level of the respective attribute, as judged by all other subjects.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

With respect to gender, the differences discussed in the main part persist. If anything, they are even more pronounced: for all memory lengths, the overall effect of a range increase if the partner is ranked above the evaluator is insignificant for male evaluators, but highly significant ($p < 0.01$) and negative for female evaluators (cf. Appendix A.1).

Result 4 (Consideration Set Composition)**.**

(a) For memory lenghts of three resp. two partners, range effects exhibit the same pattern as in the reference case of perfect memory. The effects are significantly positive when the partner ranks below the reference level ($p < 0.01$), and significantly negative when he or she ranks above ($p < 0.1$). The above reference level effect is insignificant, but directionally consistent, in case individuals are assumed to recall but one person.

(b) Differential gender effects persist when varying consideration set composition. For female evaluators, both the below and the above reference level range effects are highly significant for all memory lengths. Male evaluators conform to this pattern only partially (cf. Appendix A.1).

These results imply that context effects do not rely on a particular assumption regarding the composition of consideration sets, which in turn determines attribute ranges. Due to the fact that inferring consideration from choice data remains challenging, I adopt a mechanistic approach, in which choice objects fade from consideration over time.[22] The robustness of these results should therefore be jugded relative to the validity of the approach upon which they are based.

5 CONCLUDING REMARKS

This paper shows that choice depends on context, a notion that stands in stark contrast to the canonical *homo oeconomicus*. A fully rational agent is characterized by a set of stable preference relations on a given set of options, and adding options or taking them away does not distort preferences among the remaining ones. Exploiting a rich dataset of a speed dating experiment that contains data on both attribute utility and final choice, I provide evidence for a specific type of context dependence, first formalized by Bushong, Rabin, and Schwartzstein (2015).

Their intuition suggests that a fixed utility difference appears large when the range of possible utility values to be realized is small; but this same utility difference receives

[22] Masatlioglu, Nakajima, and Ozbay (2012), Manzini and Mariotti (2014), and Masatlioglu and Nakajima (2015) present theoretical advances regarding identification of attention and consideration. However, these approaches struggle to uncover preference completely. For the purposes of this paper, I therefore rely on a simplified framework.

less and less weight the larger the utility range becomes. In the context of speed dating, the empirical analysis supports this type of relative thinking.

Assuming for a moment that the findings of this paper are not particular to the speed dating context, but reflect a more general phenomenon, relative thinking has important policy implications. For instance, marketing specialists could exploit context dependence for strategic product placement, intentionally deemphasizing weaknesses of a product by positioning it in markets in which the utility variation along the disadvantaged dimension is large. The idea that product positioning aims at distorting consumers' perceptions of attributes is not alien to current marketing practice (cf., for instance, Boone and Kurtz 2013). The findings of this paper suggest ways in which such targets can be reached more effectively.[23]

Other potential applications include political agendas. Imagine a race between two candidates who have to decide where to locate on the political spectrum on certain issues. Suppose one candidate advocates stricter gun control, while the other prefers to maintain the more lenient status quo. For reasons of consistency, both candidates will stay true to their initial positions at least until election day. But what if the location of the median voter changes over time? This paper suggest that there are subtle ways in which candidates may influence the salience of particular issues: if public sentiment leans more towards stricter gun control, for instance, the advocate of the status quo might publicly contemplate measures that lead to more deregulation in order to make his position appear relatively less extreme.

In a similar vein, speakers of committees who have some discretion over the order in which items of the agenda are presented, could use this to their advantage. If votes are taken after each item, as is usually the case, and if items have at least partly overlapping dimensions, an agenda setter can manipulate their perceived weights. Due to the large number of possible candidates that are often evaluated along similar or identical attributes, hiring committees may be particularly prone to such effects.

Although the above examples are deliberately simplistic, they serve to show that incorporating context dependence in preferences can increase both the explanatory and predictive power of models. The central contribution of this paper therefore lies

[23]The marketing literature stresses the importance of positioning products appropriately, but is silent about ways to achieve this objective. With respect to brands, Batra, Myers, and Aaker (2009) define positioning as follows: "Just as segmentation involves the decision to aim at a certain group of customers but not others, [positioning] involves a decision to stress only certain aspects of our brand, and not others. [...]. Such positioning is achieved mostly through a brand's marketing communications, although its distribution, packaging, and actual product features can play major roles." (pp. 192–193). That product positioning aims at altering perception was first proposed by Ries and Trout (2001): "[P]ositioning is not what you do to a product. Positioning is what you do to the mind of the prospect. That is, you position the product in the mind of the prospect." (quoted from http://www.inc.com/encyclopedia/product-positioning.html).

in elucidating one particular mechanism through which context matters and in distinguishing this mechanism from other potential channels.

Nevertheless, this study has several limitations that provide fruitful areas for further research. First, the fact that decisions about potential partners are frequent (at least in terms of hypothetical evaluation), familiar, and highly consequential should make us confident that context dependence will also be apparent in less involved decisions. However, there may be other reasons why the speed dating context is idiosyncratic, and more research is needed to shed light on this issue.

Second, there appears to be a gender gap. While results are qualitatively similar for women and men, and effects point in the same direction, the coefficients on the attribute ranges are precisely estimated only for women. Whether this reflects a true gender difference or an artifact of higher heterogeneity among men remains to be shown.

Finally, the speed dating setup lacks experimental control over which potential partners are members of the consideration set, in relation to which the range is determined. Although some theoretical advances have been made in order to allow for identification of considered options (cf. Masatlioglu, Nakajima, and Ozbay 2012; Manzini and Mariotti 2014), inferring consideration from choice data is still challenging. In an attempt to alleviate related concerns, I show that results do not depend on different assumptions regarding memory length. However, seeing as the process of consideration set formation is not well understood, mechanisms other than forgetfulness might be at play.

A APPENDIX

A.1 Robustness

FULL SAMPLE The analyses presented in the main text are based on a restricted sample. As described in more detail in Section 3.2, I exclude seven of 21 sessions, two because the group size differs substantially from the average, one because a limit on positive decisions to date was in place, and four because participants were asked to bring magazines or books, which could have potentially distracted them from the task at hand.

Reassuringly, results are robust to including these omitted sessions. As shown in Table 6, coefficients point in the same direction as in Table 4, and for female participants both the coefficient on *Attractiveness Range* and the interaction with *More Attractive* are significant.

SPECIFICATION OF THE RANGE Due to the sequential nature of the speed dating context, determining the range is subject to specifying the process of consideration set formation. In the main text I assume that participants enter the experiment with no prior reference point, such that the range at the first date is unspecified.

Table 7 presents regressions that take into account one's own attribute level as reference point, relative to which the range at the first date is computed. The coefficients are not as precisely estimated, but support the analysis in the main part.

RESCALING As stated in the main text, estimated coefficients could be biased if stated attributed levels (i.e., evaluations of partners) are directly affected by context effects. To deal with this issue, Bhargava and Fisman (2014) substitute all ratings with evaluations by two research assistants seated in the same room. Applying the same methodology to context dependence faces some shortcomings: for one, it is impossible to control for attribute levels other than attractiveness, since research assistants were unable to rate these. Another concern is that ratings by research assistants might be subject to similar biases (i.e., context dependence) than those of actual participants, and hence their unbiasedness is questionable.

Nonetheless, Table 8 presents results from a stripped-down regression of the decision to date on attractiveness and the attractiveness range, as determined by average ratings of the two research assistants, including only the lagged attractiveness rating as a contrast effect control and round fixed effects to control for time trends. The results are largely in line with the findings reported in Section 4.2: for the combined sample and the female evaluator subsample, the range effect is positive (yet non-significant) if the evaluated person is below the reference level, and significantly negative if above ($F(1,399) = 4.02$, $p < 0.05$; $F(1,199) = 2.96$, $p < 0.1$). Note that the negative coef-

	Consensus Rating		
	(1) All	(2) Male	(3) Female
Attractiveness Range	0.109***	0.112***	0.0947**
	(0.0256)	(0.0324)	(0.0379)
More Attractive	0.0994***	0.0761**	0.133***
	(0.0211)	(0.0322)	(0.0297)
More Attr. × Attr. Range	-0.108***	-0.0816***	-0.163***
	(0.0178)	(0.0233)	(0.0306)
Fun Range	0.0348	0.0267	0.0403
	(0.0337)	(0.0502)	(0.0455)
More Fun	0.0668***	0.112***	0.00250
	(0.0238)	(0.0333)	(0.0328)
More Fun × Fun Range	-0.0886***	-0.106***	-0.0522
	(0.0251)	(0.0357)	(0.0350)
Constant	-0.954***	-0.994***	-0.909***
	(0.0538)	(0.0736)	(0.0774)
Attribute Level Controls	Yes	Yes	Yes
Contrast Effect Controls	Yes	Yes	Yes
Round Fixed Effects	Yes	Yes	Yes
Observations	5802	2985	2817
R^2	0.532	0.538	0.521
Adjusted R^2	0.489	0.493	0.473

Table 6.—Full Sample

Notes: Linear probability model with standard errors clustered on the partner level (in parentheses). In all regressions, the dependent variable is the Decision to Date. All observations are weighted by the inverse of the number of observations per subject; ratings are adjusted using importance weights. Attribute ratings are on a scale from 1 to 10. More Attractive and More Fun are indicator variables set to one if the partner surpasses the evaluator's own level of the respective attribute, as judged by all other subjects.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	Consensus Rating		
	(1) All	(2) Male	(3) Female
Attractiveness Range	0.127***	0.131***	0.0975*
	(0.0323)	(0.0390)	(0.0496)
More Attractive	0.133***	0.104**	0.186***
	(0.0254)	(0.0399)	(0.0342)
More Attr. × Attr. Range	-0.129***	-0.0917***	-0.227***
	(0.0220)	(0.0285)	(0.0371)
Fun Range	0.0752*	0.0243	0.110**
	(0.0438)	(0.0690)	(0.0527)
More Fun	0.0745***	0.120***	0.00444
	(0.0272)	(0.0384)	(0.0364)
More Fun × Fun Range	-0.104***	-0.117***	-0.0633*
	(0.0271)	(0.0382)	(0.0376)
Constant	-1.019***	-1.008***	-1.015***
	(0.0692)	(0.0979)	(0.0919)
Attribute Level Controls	Yes	Yes	Yes
Contrast Effect Controls	Yes	Yes	Yes
Round Fixed Effects	Yes	Yes	Yes
Observations	4710	2401	2309
R^2	0.538	0.539	0.531
Adjusted R^2	0.494	0.492	0.482

TABLE 7.—Different Starting Point for Range

Notes: Linear probability model with standard errors clustered on the partner level (in parentheses). In all regressions, the dependent variable is the Decision to Date. All observations are weighted by the inverse of the number of observations per subject; ratings are adjusted using importance weights. Attribute ratings are on a scale from 1 to 10. More Attractive and More Fun are indicator variables set to one if the partner surpasses the evaluator's own level of the respective attribute, as judged by all other subjects.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	Consensus Rating		
	(1) All	(2) Male	(3) Female
Attractiveness Range	0.0424	-0.0190	0.141
	(0.0518)	(0.0613)	(0.107)
More Attractive	0.266***	0.247***	0.315***
	(0.0262)	(0.0385)	(0.0377)
More Attr. × Attr. Range	-0.149***	-0.0894**	-0.321***
	(0.0324)	(0.0399)	(0.0618)
Constant	-0.363***	-0.317***	-0.442***
	(0.0518)	(0.0749)	(0.0719)
Attribute Level Controls	Yes	Yes	Yes
Contrast Effect Controls	Yes	Yes	Yes
Round Fixed Effects	Yes	Yes	Yes
Observations	5656	2831	2825
R^2	0.401	0.399	0.395
Adjusted R^2	0.354	0.349	0.344

TABLE 8.—Rescaling (Research Assistant Rating)

Notes: Linear probability model with standard errors clustered on the partner level (in parentheses). In all regressions, the dependent variable is the Decision to Date. All observations are weighted by the inverse of the number of observations per subject; ratings are adjusted using importance weights. Attribute ratings are on a scale from 1 to 10; all ratings are replaced by the average of two research assistants' ratings. More Attractive is an indicator variable set to one if the partner surpasses the evaluator's own level of this attribute, as judged by all other subjects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

ficient of the range on the decision to date in case the partner is above the reference level is also present for male evaluators ($F(1, 199) = 2.85$, $p < 0.1$).

As an additional robustness check, I propose a different strategy, which rests on the assumption that evaluators cannot be influenced by context effects in the very first round. In this case we can take the assessment of the evaluator who has dated the target in the first round, and apply this rating to assessments of that target in all subsequent rounds.

The plausibility of this assumption strongly depends on whether participants enter the speed dating experiment with some expectations regarding reference attribute levels or not. If they do, they might evaluate even the first target relative to their expectation, and hence be influenced by context effects even then. However, if reference attribute levels are not present or carry sufficiently low weight, we can reasonably expect that the first rating is unimpeded by distortions due to context effects. In this case, it would be natural to take the assessment of the evaluator who has dated the target in the first round, and apply this rating to assessments of that target in all subsequent rounds.

One obvious drawback of this method relative to the one employed by Bhargava and Fisman (2014) is that 'true' attribute levels are likely to be measured less precisely, as they rely on the rating of one instead of two persons. As pointed out before, however, context effects may be present in research assistants' ratings, whereas they are very unlikely if we employ the rating of that evaluator who has dated a given target in the first round. An additional advantage of the method I propose is that it yields undistorted ratings for all attributes, not just attractiveness.

Substituting all ratings with those of the subject who dated a given person first yields Table 9. Results are qualitatively in line with Table 4, but are not significant. However, considering that there is a substantial variation in the perception of attributes across subjects, which this method fails to capture, this loss in precision is expected.

CONSIDERATION SET COMPOSITION This section complements the analysis in Section 4.4. In order to evaluate whether there are differential effects of consideration set composition on context dependence, Tables 10 and 11 split the sample into male and female evaluators.

A.2 *Gender Differences in Utility*

The results presented in Section 4.2 suggest that the effect of utility ranges is stronger for women than for men, and loses significance when considering only the data generated by male evaluators. One possible explanation for this finding is that men do not care about how attractive their counterparts are, implying that utility is flat over all levels of attractiveness. In this case the proposed identification strategy would not work, since it is based on differential reactions to range increases depending on whether the level of attractiveness is regarded as desirable or repelling.

Figure 6 documents that this rationalization does not hold: if anything, the positive slope of attractiveness utility is even steeper than for women. In addition, note that for both women and men, the consensus rating approximates the utility threshold very well.

	Consensus Rating		
	(1) All	(2) Male	(3) Female
Attractiveness Range	-0.00475	0.0000181	0.0199
	(0.0294)	(0.0330)	(0.0579)
More Attractive	0.231***	0.253***	0.252***
	(0.0295)	(0.0395)	(0.0459)
More Attr. × Attr. Range	-0.0325	-0.0314	-0.0927**
	(0.0215)	(0.0251)	(0.0444)
Fun Range	-0.00383	-0.0315	-0.0106
	(0.0543)	(0.0713)	(0.0787)
More Fun	0.234***	0.334***	0.0755
	(0.0410)	(0.0531)	(0.0622)
More Fun × Fun Range	-0.0731*	-0.150***	0.0630
	(0.0381)	(0.0469)	(0.0606)
Constant	0.0187	-0.0446	0.104
	(0.0938)	(0.131)	(0.136)
Attribute Level Controls	Yes	Yes	Yes
Contrast Effect Controls	Yes	Yes	Yes
Round Fixed Effects	Yes	Yes	Yes
Observations	4447	2239	2208
R^2	0.385	0.396	0.376
Adjusted R^2	0.320	0.328	0.304

Table 9.—Rescaling (First Rating)

Notes: Linear probability model with standard errors clustered on the partner level (in parentheses). In all regressions, the dependent variable is the Decision to Date. All observations are weighted by the inverse of the number of observations per subject; ratings are adjusted using importance weights. Attribute ratings are on a scale from 1 to 10; all ratings are replaced by the assessment of the evaluator who has rated the partner in the first round. More Attractive and More Fun are indicator variables set to one if the partner surpasses the evaluator's own level of the respective attribute, as judged by all other subjects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	Consensus Rating			
	(1) Perf. Mem.	(2) 3 Partners	(3) 2 Partners	(4) 1 Partner
Attractiveness Range	0.116***	0.0870***	0.0953***	0.130***
	(0.0350)	(0.0304)	(0.0318)	(0.0376)
More Attractive	0.0963***	0.0704**	0.0652*	0.0545*
	(0.0369)	(0.0350)	(0.0341)	(0.0292)
More Attr. × Attr. Range	-0.0900***	-0.0858***	-0.0941**	-0.113**
	(0.0269)	(0.0320)	(0.0363)	(0.0466)
Fun Range	0.0279	0.0996**	0.0715*	0.0781
	(0.0551)	(0.0441)	(0.0423)	(0.0511)
More Fun	0.110***	0.0990***	0.0888***	0.0769**
	(0.0370)	(0.0362)	(0.0341)	(0.0315)
More Fun × Fun Range	-0.112***	-0.132***	-0.130**	-0.162**
	(0.0392)	(0.0498)	(0.0556)	(0.0745)
Constant	-0.967***	-0.930***	-0.901***	-0.886***
	(0.0838)	(0.0855)	(0.0842)	(0.0808)
Attribute Level Controls	Yes	Yes	Yes	Yes
Contrast Effect Controls	Yes	Yes	Yes	Yes
Order Fixed Effects	Yes	Yes	Yes	Yes
Observations	2401	2401	2401	2401
R^2	0.539	0.538	0.538	0.541
Adjusted R^2	0.492	0.491	0.491	0.494

Table 10.—Consideration Set Composition (Men)

Notes: Linear probability model with standard errors clustered on the partner level (in parentheses). In all regressions, the dependent variable is the Decision to Date. All observations are weighted by the inverse of the number of observations per subject; ratings are adjusted using importance weights. Attribute ratings are on a scale from 1 to 10. More Attractive and More Fun are indicator variables set to one if the partner surpasses the evaluator's own level of the respective attribute, as judged by all other subjects.

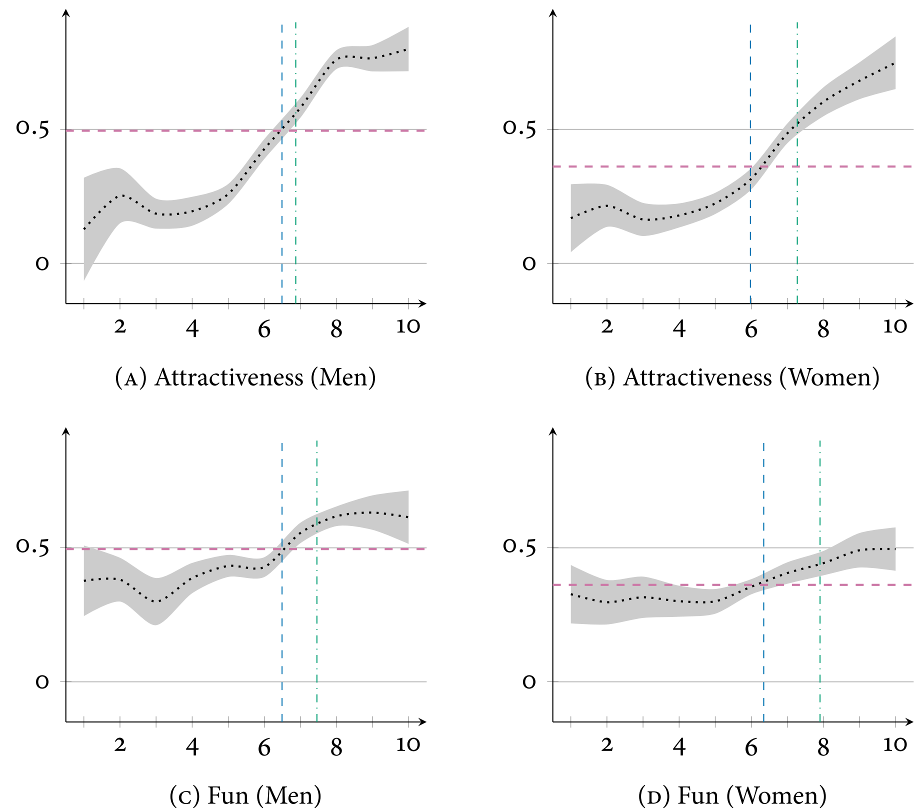
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	Consensus Rating			
	(1) Perf. Mem.	(2) 3 Partners	(3) 2 Partners	(4) 1 Partner
Attractiveness Range	0.112***	0.0825**	0.0912***	0.170***
	(0.0415)	(0.0389)	(0.0331)	(0.0407)
More Attractive	0.176***	0.134***	0.120***	0.113***
	(0.0323)	(0.0284)	(0.0282)	(0.0262)
More Attr. × Attr. Range	-0.226***	-0.218***	-0.227***	-0.347***
	(0.0357)	(0.0411)	(0.0462)	(0.0668)
Fun Range	0.100**	0.147***	0.114***	0.103***
	(0.0491)	(0.0388)	(0.0356)	(0.0359)
More Fun	-0.00420	-0.00380	-0.0320	-0.0240
	(0.0355)	(0.0295)	(0.0283)	(0.0275)
More Fun × Fun Range	-0.0556	-0.0865*	-0.0441	-0.0862
	(0.0390)	(0.0439)	(0.0472)	(0.0646)
Constant	-0.975***	-0.962***	-0.920***	-0.867***
	(0.0840)	(0.0799)	(0.0782)	(0.0751)
Attribute Level Controls	Yes	Yes	Yes	Yes
Contrast Effect Controls	Yes	Yes	Yes	Yes
Order Fixed Effects	Yes	Yes	Yes	Yes
Observations	2309	2309	2309	2309
R^2	0.531	0.529	0.527	0.529
Adjusted R^2	0.482	0.479	0.477	0.479

TABLE 11.—Consideration Set Composition (Women)

Notes: Linear probability model with standard errors clustered on the partner level (in parentheses). In all regressions, the dependent variable is the Decision to Date. All observations are weighted by the inverse of the number of observations per subject; ratings are adjusted using importance weights. Attribute ratings are on a scale from 1 to 10. More Attractive and More Fun are indicator variables set to one if the partner surpasses the evaluator's own level of the respective attribute, as judged by all other subjects.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$



(A) Attractiveness (Men) (B) Attractiveness (Women)

(C) Fun (Men) (D) Fun (Women)

FIGURE 6.—Gender Differences in Utility

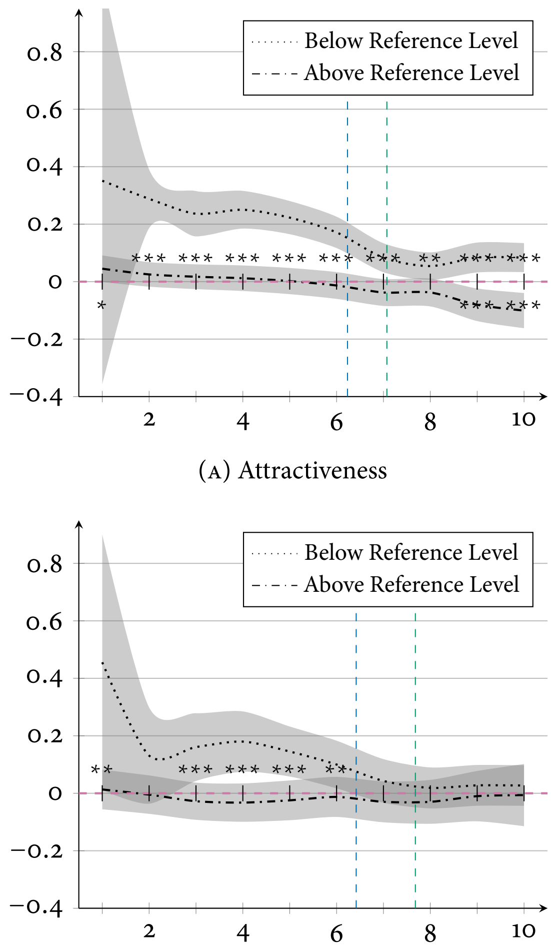
Notes: Vertical axis displays average predicted values of the dependent variable (decision to date) for different attribute values. Plotted values are derived from regressing the decision to date on all five attributes, including time and individual fixed effects. Marginal effects are obtained post estimation for each attribute separately. Regressions are weighted by the number of observations in each session. Standard errors are clustered on the partner level. Grey areas are 95% confidence intervals. Horizontal red lines depict average yes rate, vertical blue lines (dashed) are average consensus ratings of the respective attribute, and vertical green lines (dash-dotted) are average own ratings of the attribute.

A.3 Uniqueness of the Reference Level

This section complements the analysis presented in Section 4.3 in the main part.

Figure 7 displays the results for placebo regressions performed on the combined sample. As discussed in the main part, if there were indeed a unique reference level of utility, we would expect to find a significant *below* reference level range coefficient for values up to some reference level, a significant *above* reference level coefficient for values above, and a null effect for all other levels of these two coefficients. Panel 7a shows a pattern that is qualitatively consistent with this expectation, but fails to hold up quantitatively. In particular, there is no apparent change in both the magnitude and the significance of the coefficients at some value between the consensus rating and the own rating. Furthermore, the below threshold coefficient remains significantly positive even for levels above the threshold. Recalling that there appears to be no generalized context dependence across genders (cf. Table 4), this finding is expected.

Results for the attribute *fun* are less clear-cut. Some of the below reference level estimates are consistent with the logic outlined above: namely, that there is a negative range effect for placebo reference level values up to approximately 6, and that the effect is indistinguishable from zero above. However, the above reference level estimates do not conform to this logic. This pattern is consistent with the insignificant findings reported in Result 2: it appears that there is no unique reference level of fun, but that there are many. Hence we cannot expect to find meaningful results when taking the consensus rating as a benchmark.

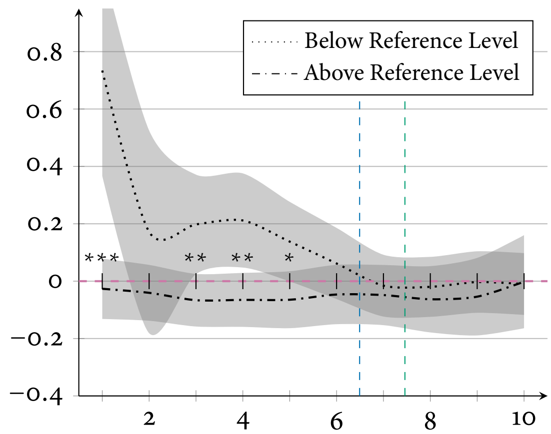


(A) Attractiveness

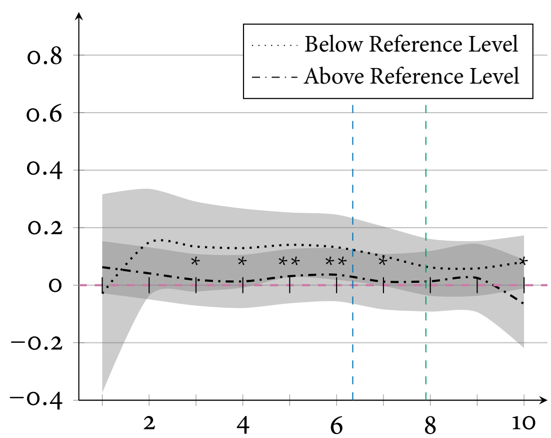
(B) Fun

FIGURE 7.—Placebo Tests of Attribute Ranges

Notes: Figure displays regression coefficients of the decision to date on attribute range (dotted line) and the sum of attribute range and the interaction between attribute range and a dummy (dash-dotted line). The dummy is set to one if the attribute level is higher than $k = 1, 2, \ldots, 10$ and set to zero otherwise, such that each line consists of coefficients from 10 separate regressions. Grey areas are 95% confidence intervals. The horizontal axis tracks the value of the threshold k. The two vertical lines indicate the averages of the consensus rating of the attribute (blue line) respectively subject's own rating of the attribute (green line). Stars above the red line indicate whether coefficients of the dotted line are significantly different from zero; analogously for stars below the red line and the dash-dotted line. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.



(A) Men



(B) Women

FIGURE 8.—Placebo Tests of Fun Ranges

Notes: Figure displays regression coefficients of the decision to date on attribute range (dotted line) and the sum of attribute range and the interaction between attribute range and a dummy (dash-dotted line). The dummy is set to one if the attribute level is higher than $k = 1, 2, \ldots, 10$ and set to zero otherwise, such that each line consists of coefficients from 10 separate regressions. Grey areas are 95% confidence intervals. The horizontal axis tracks the value of the threshold k. The two vertical lines indicate the averages of the consensus rating of the attribute (blue line) respectively subject's own rating of the attribute (green line). Stars above the red line indicate whether coefficients of the dotted line are significantly different from zero; analogously for stars below the red line and the dash-dotted line. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

REFERENCES

Batra, Rajeev, John G. Myers, and David A. Aaker. 2009. *Advertising Management*. London: Dorling Kindersley.

Bhargava, Saurabh, and Raymond Fisman. 2014. "Contrast Effects in Sequential Decisions: Evidence from Speed Dating." *Review of Economics and Statistics* 96 (3): 444–457.

Boone, Louis E., and David L. Kurtz. 2013. *Contemporary Marketing*. Boston: Cengage Learning.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2013a. "Salience and asset prices." *The American Economic Review* 103 (3): 623–628.

— . 2013b. "Salience and Consumer Choice." *Journal of Political Economy* 121 (5): 803–843.

— . 2012a. "Salience in experimental tests of the endowment effect." *The American Economic Review* 102 (3): 47–52.

— . 2012b. "Salience Theory of Choice Under Risk." *The Quarterly Journal of Economics* 127 (3): 1243–1285.

— . 2015. "Salience theory of judicial decisions." *The Journal of Legal Studies* 44 (S1): S7–S33.

Bracha, Anat, Uri Gneezy, and George Loewenstein. 2015. "Relative Pay and Labor Supply." *Journal of Labor Economics* 33 (2): 297–315.

Bushong, Benjamin, Matthew Rabin, and Joshua Schwartzstein. 2015. "A Model of Relative Thinking." *mimeo*.

Dertwinkel-Kalt, Markus, et al. 2016. "Concentration Bias in Intertemporal Choice." *mimeo*.

Fisman, Raymond, et al. 2006. "Gender Differences in Mate Selection: Evidence From a Speed Dating Experiment." *The Quarterly Journal of Economics* 121 (2): 673–697.

— . 2008. "Racial Preferences in Dating." *The Review of Economic Studies* 75 (1): 117–132.

Huffman, David, and Matias Barenstein. 2005. "A Monthly Struggle for Self-Control? Hyperbolic Discounting, Mental Accounting, and the Fall in Consumption Between Paydays." *Institute for the Study of Labor (IZA) Discussion Paper 1430*.

Kőszegi, Botond, and Matthew Rabin. 2006. "A Model of Reference-Dependent Preferences." *The Quarterly Journal of Economics* 121 (4): 1133–1165.

— . 2007. "Reference-Dependent Risk Attitudes." *The American Economic Review* 97 (4): 1047–1073.

Kőszegi, Botond, and Adam Szeidl. 2013. "A Model of Focusing in Economic Choice." *The Quarterly Journal of Economics* 128 (1): 53–104.

Manzini, Paola, and Marco Mariotti. 2014. "Stochastic Choice and Consideration Sets." *Econometrica* 82 (3): 1153–1176.

Masatlioglu, Yusufcan, and Daisuke Nakajima. 2015. "Completing Incomplete Revealed Preference Under Limited Attention." *Japanese Economic Review* 66 (3): 285–299.

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y. Ozbay. 2012. "Revealed Attention." *The American Economic Review* 102 (5): 2183–2205.

Niederle, Muriel. 2016. "Gender." In *Handbook of Experimental Economics, Vol. 2*, ed. by John H. Kagel and Alvin E. Roth. Princeton: Princeton University Press.

Pepitone, Albert, and Mark DiNubile. 1976. "Contrast Effects in Judgments of Crime Severity and the Punishment of Criminal Violators." *Journal of Personality and Social Psychology* 33 (4): 448–459.

Ries, Al, and Jack Trout. 2001. *Positioning: The Battle for Your Mind*. New York City: McGraw-Hill.

Soltani, Alireza, Benedetto De Martino, and Colin Camerer. 2012. "A Range-Normalization Model of Context-Dependent Choice: A New Model and Evidence." *PLoS Computational Biology* 8 (7): e1002607.

Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (4481): 453–458.