MORALS, MARKETS, AND MALLEABILITY: THREE ESSAYS IN BEHAVIORAL ECONOMICS

MARIUS-FLORENTIN KRÄMER

Dissertation

Marius-Florentin Krämer: *Morals, Markets, and Malleability: Three Essays in Behavioral Economics,* Dissertation, December 5, 2016

MORALS, MARKETS, AND MALLEABILITY: THREE ESSAYS IN BEHAVIORAL ECONOMICS

INAUGURAL-DISSERTATION

zur Erlangung des Grades

DOCTOR OECONOMIAE PUBLICAE (DR. OEC. PUBL.)

an der LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

2016

vorgelegt von MARIUS-FLORENTIN KRÄMER

Referent:	PROF. DR. KLAUS M. SCHMIDT
Korreferent:	PROF. DR. MARTIN G. KOCHER
Promotionsabschlußberatung:	10. MAI 2017
Berichterstatter:	PROF. DR. KLAUS M. SCHMIDT
	PROF. DR. MARTIN G. KOCHER
	PROF. DR. FABIAN HERWEG
Mündliche Prüfung:	04. MAI 2017

ACKNOWLEDGMENTS

The present dissertation is the result of inspiring discussions, extensive readings, stimulating summer schools, workshops and conferences as well as continuous questioning of what seems to be obvious. I am deeply indebted to my first supervisor Klaus Schmidt for creating an environment that enabled me to pursue my research interests with just the right amounts of independence and guidance. Beyond offering extremely valuable comments and suggestions for all research projects that are part of this dissertation (and those that he wisely advised me not to pursue), he taught me how to think analytically, how to write clearly and purposefully, and how to find out whether a question is truly interesting or merely unanswered.

Florian Englmaier, Joachim Winter and all other faculty members associated with the graduate program "Evidence-Based Economics" deserve a very special thank you. Not only did they design a fantastic course program with highly relevant core courses and top-class, inspiring guest lecturers; they also created an open, stimulating environment that encouraged me to think big (even though that sometimes led into dead ends). The graduate program defines itself via a common methodological framework, as opposed to commonality of subject matter; this philosophy is reflected in the heterogeneity of research interests of its participants. As a result, discussions tended to shed light on aspects from a variety of angles, contributing significantly to both breadth and depth of understanding.

The first two chapters of this dissertation are joint projects. It was always pleasurable, interesting and inspiring to work with Klaus Schmidt, Martin Spann and Lucas Stich, and I will remember the lively discussions we had. Lucas has become a good friend over the years, and I am grateful to him for making work that much more enjoyable and fun.

During the last couple of years existing friendships have deepened and new ones have formed. David Schindler and Mark Westcott have liberally offered encouragement, feedback, philosophic debates and welcome distractions from the day to day business. Johannes Maier has been indispensable for questioning new ideas and helping me see to the core of their implications (although this often resulted in their eventual abandon). Fabian Herweg, Clemens König and Peter Schwardmann always had open ears and offered valuable advice during all stages of

my time at LMU. Finally, I would like to thank all my colleagues at the chairs of Klaus Schmidt and Martin Kocher for creating an atmosphere in which both work and leisure had appropriate value.

During the past year, I have been employed as the lab manager of MELESSA. As director of the lab, Martin Kocher was always helpful and very responsive in case questions came up, and I enjoyed the casual, yet efficient and professional work atmosphere at the lab. Naturally, running experiments well is an art as much as it is science, and I am glad to say that the student research assistants at MELESSA have brought this art to perfection.

Most of all, I would like to express my deep and heartfelt gratitude to my parents, Barbara and Steffen, my brothers and sisters, Vivika, Magali, Felicia and Laurin, my friend Lukas, and my wife Caroline. Their constant support, encouragement, and occasional ridicule helped me to stay focused without taking myself too seriously.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Part I

PREFACE

PREFACE

"We absolutely must leave room for doubt or there is no progress and there is no learning. There is no learning without having to pose a question. And a question requires doubts. People search for certainty. But there is no certainty. People are terrified—how can you live and not know? It is not odd at all. You only think you know, as a matter of fact. And most of your actions are based on incomplete knowledge and you don't really know what it is all about, or what the purpose of the world is, or know a great deal of other things. It is possible to live and not know."

— Richard Feynman, 1964

Economics is about striving to understand human behavior in situations of economic interest, and so it seems almost redundant to speak of *Behavioral Economics* as a distinct discipline. Yet it has emerged as a subfield of economics that keeps a certain distance from neoclassical thinking. Perhaps this distance is best explained by the fact that *decriptively* modeling human behavior with all its biases, inconsistencies, and irrational tendencies often results in intractability and lack of generality. The alternative approach, in which some *normatively* desirable qualities are presented, and outcomes of decisions are measured by appealing to this ideal, enables relatively accurate analyses and predictions under many circumstances, while maintaining parsimony. In addition, neoclassical economics lends itself to conducting welfare analysis by insisting that preferences can be recovered from choice data; once we allow for "mistakes", such analysis becomes more difficult (but not impossible, as shown for instance by Kőszegi and Rabin (2008)).

There is hence a tension between parsimony and realism. Parsimony entails the pursuit of elegance, the quest for maximal power at minimal structure. Realism, on the other hand, seeks to make precise certain aspects of behavior, which sometimes comes at the cost of loss in generality. Compromising between these two *desiderata* is among the most important functions a model can hope to accomplish. Striking the right balance results in models that aspire to be as general as possible and as realistic as necessary. Put differently, realism is useful not for its own

sake, but only insofar as it contributes to our understanding of economic phenomena.[1]

Whether it is necessary to diverge from the normatively desirable assumptions on behavior imposed by the neoclassical school therefore depends on the objective and scope of the question one hopes to answer. In some areas, it is sufficient to rely on traditional behavioral assumptions in order to characterize and predict economic outcomes; in others, taking into account systematic behavioral variations matters a great deal. To name but a few, it appears that humans do not always pursue their narrow self-interest (E. Fehr & Fischbacher, 2002), make time-inconsistent choices (Frederick, Loewenstein, & O'Donoghue, 2002), incorrectly infer population probabilities from observing only a small sample (the "law of small numbers", cf. Rabin (2002)), and are sensitive to outcomes that could have been realized, anticipating that they might regret their (*ex ante* correct) choice *ex post* (Bell, 1982; Loomes & Sugden, 1982; Filiz-Ozbay & Ozbay, 2007).

All in all, I believe that the question ultimately is not *whether* to model behavior, but *which* aspects of it. To judge whether it is necessary to incorporate some systematic behavioral pattern, the explanatory and predictive power of behaviorally enriched (relative to standard) models must serve as guiding principle. To speak of *Behavioral Economics* as a somewhat disctinct sub-discipline may therefore be misleading: economics is behavioral *per se*. Yet despite the tautological nature of this term, it has come to refer to a general acknowledgement of the need to incorporate into our models systematic behavioral patterns that are relevant to questions of economic interest. The subtitle of this dissertation, *Three Essays in Behavioral Economics*, reflects this line of thought.

It remains to flesh out the three terms I use to characterize the chapters of this dissertation: *Morals, Markets, and Malleability*. Chapter 1 is concerned with a century-old philosophical debate about the limits of ethics. It has long been unclear whether animals have a place in ethical frameworks, which are mostly concerned with interaction among *humans*. Immanuel Kant, for instance, denies animals intrinsic rights to be treated well, but argues that treatment of animals will reflect onto conduct among humans, thereby proposing a consequentialist argument for including animals in ethical considerations. By asking whether individuals who have a higher propensity for prosocial behavior towards

1 See Schmidt (2009) for a much more detailed overview of the interaction between economic experiments, which are designed to study behavior descriptively, and the development of models and theory.

other humans are willing to pay more for animal welfare, we contribute to this debate empirically.[2]

Chapter 2 is inspired by novel pricing formats that have emerged in recent years. Instead of setting a fixed price, these mechanisms delegate some pricing power to customers. We are interested in the performance of Pay What You Want and Name Your Own Price vis-à-vis traditional posted price sellers. To gain insights into the behavior of customers and sellers facing these innovative pricing mechanisms, we conduct an experiment that allows us to vary market structure, valuations, and costs exogenously.

Finally, Chapter 3 addresses a line of research that has been met with growing interest by psychology and economics scholars alike. By assuming that preferences are malleable and may be shaped by context, this research agenda simultaneously repudiates one of the central assumptions of neoclassical economics and opens new avenues for explaining heretofore puzzling behavior. Yet despite the wealth of theoretical models that aim to capture context dependence in preferences, little is known from an empirical perspective. In this chapter, I use a speed dating experiment to directly compare the central assumptions of two models, allowing me to identify one possible mechanism for context dependence in a natural setting.

Disparate though they may seem, the three chapters of this thesis are connected by a common thread. On a very broad level, each acts on the assumption that behavior may not be rational in the canonical sense. Individuals appear to be willing to pay substantial amounts for animal welfare, offer positive amounts of money to completely anonymous experimental subjects if they can get the same good without paying anything, and are sensitive to the context in which a decision is made.

All three findings violate the tenets of neoclassical economics, which is built on self-interest and stable preferences. The first two chapters show that behavior is not solely driven by selfish interests. Instead, it seems that

2 The fact that Kant proposes a *consequentialist* argument that aims at improving animal welfare deserves special notice, considering that his *Categorical Imperative* is firmly rooted in a deontological framework. One way of rationalizing this statement is that at the time, animals were indeed not endowed with intrinsic moral rights, and better treatment emerged through utilitarian behavior rather than acceptance of some universal norm. In recent work, Falk and Tirole (2016) approach the tension between deontological and consequentialist ethics by introducing narratives and imperatives into a utilitarian framework. They show that under certain conditions, a principal interested in persuading an agent to act in moral ways will use imperatives to do so. Deontological behavior thus emerges endogenously, even though all agents are consequentialists.

other humans, and even non-human animals, can be part of individuals' utility functions. The third chapter addresses a line of research that has received a lot of attention in recent years and is connected by the stipulation that preferences may not be stable across choice contexts.[3] I contribute to this literature by providing an empirical test of context dependence and showing that in the area of speed dating, decision-making is guided by relative thinking, as conceptualized by Bushong et al. (2015).

On a more narrow level, the three chapters are connected by a common methodological approach. Learning about behavior requires tools. Traditional empirical methods and econometric techniques, using observational data, have been complemented by controlled experiments.[4] Experiments are extremely useful to isolate factors that causally impact behavior, and as such they are intimately linked to the development of *Behavioral Economics* as a sub-discipline.

ANIMAL WELFARE AND HUMAN ETHICS The first chapter of this dissertation, which is joint work with Konstanze Albrecht and Nora Szech, describes an experiment aimed at eliciting participants' willingness to pay for animal welfare. We argue that in order to measure concern for animal welfare, it is informative to use a controlled experimental setup rather than questionnaire studies or purchasing data. Questionnaire studies have often been criticized for measuring intent rather than behavior, and indeed it seems that this phenomenon is quite large in magnitude when it comes to ethically produced food (cf. Bray, Johns, & Kilburn, 2010). Purchasing data do not contain information about underlying reasons for behavior. Since demand for organic food may be partly determined by advertising, image or health concerns, such data do not suffice to gain insights about concern for animal welfare.

3 This literature extends from models of rational inattention (Sims, 2003, 2006; Matejka & McKay, 2014) and limited attention (Gabaix & Laibson, 2006; Masatlioglu, Nakajima, & Ozbay, 2012; Manzini & Mariotti, 2014) to applications of these ideas to public policy and industrial organization contexts (Chetty, Looney, & Kroft, 2009; Heidhues, Kőszegi, & Murooka, 2016, forthcoming). This class of models also encompasses theories of salience (Bordalo, Gennaioli, & Shleifer, 2012, 2013; Kőszegi & Szeidl, 2013; Bushong, Rabin, & Schwartzstein, 2015), which show how decision weights may be distorted and thus lead to over- or underweighting of some choice dimensions relative to others.

4 A classification of the different types of experiments is beyond the scope of this dissertation. The methodological framework underlying laboratory experiments is summarized in Smith (2008). For excellent reviews of field experiments see Harrison and List (2004) and DellaVigna (2009).

To close this gap in the literature and to better understand what motivates consumers to purchase organic foods, we elicit concern for animal welfare in an *incentivized, direct* and *real* setup, in a task that allows to abstract from self-oriented motives. Subjects choose between intensive farming and organic farming for a living laying hen. Opting for organic farming is costly, yet guarantees better food, daylight, and much more space to the hen. This procedure allows us to elicit subjects' willingness to pay for transferring a hen to an environment that is better suited to its needs. Full anonymity and a single-blind experimental design ensure that our results are not diluted by image concerns or experimenter demand effects. We can therefore isolate concern for organic agriculture and animal welfare from other more self-oriented factors that could be at play when purchasing organic food at a supermarket.

We find that subjects are willing to pay 14 Euros for animal welfare on average. They exhibit substantial heterogeneity: while 37% are willing to pay at least 25 Euros in order to put a laying hen into better living conditions, 15% are not willing to pay anything.

We also investigate a question that has received a lot of attention by philosophers and ethicists, starting with Thomas Aquinas in the 13th century. In its most general form, this question concerns the limits of ethics, i.e. where to draw boundaries between subjects that should or should not be part of a moral universe. Non-human animals have traditionally been denied the intrinsic right to be treated well; in a deontological framework, they would therefore be outside the scope of ethical considerations. However, philosophers like Kant, Locke and Descartes agreed that treating animals well might result in more civilized conduct among *humans*. To be more precise, Kant holds the view that treating animals well is a means to an end, but not an end in itself: "If he is not to stifle his human feelings, he must practice kindness towards animals, for he who is cruel to animals becomes hard also in his dealings with men" (Gruen, 2014).

Kant thus claims that there is a link between the treatment of animals and concern for fellow human beings. This relationship is key for the central question we aim to answer in this chapter: rather than contributing to the debate about environmental consequences and sustainability of intensive animal farming, we are concerned with its *social* importance. By eliciting prosocial motivations using validated psychological scales, and combining these with the willingness to pay for animal welfare, our experiment sheds light on the correlation between concern for animal welfare and respect towards other humans. The results suggest that this link exists: subjects who score higher on the tactics dimension of the

Machiavellianism scale, meaning they are more willing to circumvent conventional moral standards in order to achieve personal gains, exhibit a lower willingness to pay for animal welfare.

DELEGATING PRICING POWER TO CUSTOMERS The second chapter of this dissertation is joint work with Klaus Schmidt, Martin Spann and Lucas Stich. It is based on the observation that some firms increasingly use novel types of pricing instruments that delegate price-setting power to customers, either in part or in full. Full delegation occurs when customers are at liberty to pay as much as they would like, a mechanism that has become known as "Pay What You Want" (PWYW). A slightly different pricing format lets sellers set a secret reserve price, against which potential customers can submit a bid. If the bid is higher than the reserve price, the transaction takes place at the price of the bid; otherwise neither party receives a payoff. Since buyers have some discretion over the price by determining the amount they would like to bid, delegation of pricing power is only partial. This mechanism is commonly known as "Name Your Own Price" (NYOP).

The contribution of this chapter is twofold: first, we document how PWYW and NYOP sellers perform relative to posted-price sellers within a unified experimental paradigm. This is in contrast to the previous literature, which studies these mechanisms in isolation. We also explore the behavior of customers, which enables us to speak to the differences we observe across the different market structures and pricing mechanisms. Second, we explicitly allow for direct and indirect promotional benefits that may occur as a consequence of using these novel types of pricing formats. Their participatory and innovative nature often initiates word-of-mouth recommendations and favorable press coverage (Hinz & Spann, 2008; Kim, Natter, & Spann, 2009). Thus, a direct benefit is that they can be a powerful tool to promote a product to a wider audience. In addition, PWYW and NYOP may result in indirect promotional benefits by increasing demand for complementary products.

Our results indicate that both PWYW and NYOP can be profitable alternatives to posted prices. However, their scope of application is very different: PWYW is very successful in terms of market penetration, but it fails to generate sustainable profits in the absence of (exogenously generated) promotional benefits. While a substantial fraction of experimental subjects are willing to pay strictly positive amounts, and in some cases even are prepared to compensate sellers for their costs, average payments are not sufficient to keep PWYW sellers in business. Anticipating this,

PWYW sellers either do not enter the market or elect to use posted prices in case the promotional benefit does not exceed their costs.

This picture changes when we take into account promotional benefits that may result from using such an innovative mechanism. Conditional on entering the market and choosing the PWYW mechanism, a seller generates high profits, while at the same time driving competing posted-price sellers out of the market. This observation hints at another function of PWYW, *predatory pricing*: by avoiding setting a fixed price, PWYW sellers create incentives for buyers to undercut competitors' prices. Yet such a pricing strategy does not run the risk of lowering the reference price to zero (as would be the case if the product was given away for free). This suggests that PWYW can be profitably employed as a short-term tool to harm existing competitors or to deter entry of new firms.

However, it is debatable whether promotional benefits exist, and if so, whether they are sufficient to offset potential losses in revenue that may result from employing customer-driven pricing mechanisms. In the first set of experiments, we presume these benefits into existence by pointing out several channels through which they could materialize. This approach has two main advantages: first, it can help us understand how customer-driven pricing mechanisms perform if we account for benefits in a very general sense, i.e. without specifying their exact source. This should make us immune against claims that the experimental design fails to capture certain promotional benefit channels. Second, exogenously imposing promotional benefits reduces the complexity of the experiment considerably, thereby diminishing the potential for errors due to incomplete understanding.

While a tractable and general approach in principle, this first set of experiments cannot alleviate concerns related to the *existence* of promotional benefits. This is why we report on a second set of experiments, in which we allow for benefits to arise endogenously by implementing one potential channel, word of mouth. In order to do so, we split buyers into a well-informed and a follower segment. Well-informed buyers have complete information over the market structure, whereas follow-up buyers have to rely on word of mouth in order to see which sellers are present. We show that in this setting, PWYW continues to monopolize the well-informed segment of the market, which then leads to monopolization of the entire market, since follow-up buyers hardly ever get to see competing sellers. Despite almost complete market penetration, PWYW sellers do not generate profits on the follow-up market, an indication that this particular channel cannot account for promotional benefits in the case of PWYW. Turning our focus to NYOP sellers, a different picture

emerges: their market share on the follow-up market is lower, but still twice as high as that of the competing posted-price sellers. In contrast to PWYW sellers, NYOP sellers are able to translate this increased sales volume into profits that are considerably higher than those of competing sellers. The endogenous emergence of promotional benefits due to word of mouth suggests that this is an important determinant of the profitability of NYOP.

Going back to the first set of experiments, in which it was randomly determined whether benefits were present or not, we see that NYOP is profitable even in the absence of direct or indirect promotional benefits. In fact, about 60% of buyers choose the NYOP seller over a posted price competitor. Transaction prices are substantial and cover sellers' costs at all cost-benefit levels. In contrast to PWYW, buyers' degree of prosociality has no effect on submitted bids. NYOP is much less aggressive in terms of its competitive effects; it leaves room for competing posted price sellers and relaxes competition.

In conclusion, PWYW and NYOP are pricing mechanisms that delegate pricing power to customers, albeit to a different extent. Despite their upfront similarities, they serve different objectives: PWYW is very successful at penetrating the market, which can be advantageous if capacity constraints are negligible, marginal costs are low, and sellers expect to profit from spillover effects on complementary products. NYOP, on the other hand, insures the seller against losses even at high cost levels, making this a viable alternative to posted prices over extended periods of time. When being offered a choice between purchasing the same good at a posted-price seller or a NYOP seller, high-valuation customers shy away from the risk associated with submitting a potentially unsuccessful bid at the NYOP seller. In combination with the fact that the NYOP mechanism refrains from setting a publicly observable price, this segmentation can also help to avert harm from brands that are positioned in high price segments by allowing them to sell excess capacity via third party intermediaries.

CONTEXT DEPENDENCE IN SPEED DATING In the third chapter of this dissertation I take up an old question that has been met with renewed interest in the recent economic literature: do preferences depend on context, and if so, how? Cognitive psychologists propose that *perception* depends on context, meaning that objects may be perceived differently depending on their surroundings. The question I aim to answer in this chapter goes one step further by asking whether and how such perceptive errors translate into *behavior*.

To study this question I use data from a speed dating experiment first described in Fisman, Iyengar, Kamenica, and Simonson (2006). I argue that this experiment provides an ideal setup to test for a specific type of context dependence underlying the theoretical models by Kőszegi and Szeidl (2013) and Bushong et al. (2015). In these frameworks, the range of utility levels along a given attribute determines whether this attribute will be over- or underweighted. In order to provide a direct empirical test of context dependence, one would ideally need a dataset that contains information about attribute utility values and exhibits exogenous variation in attribute utility ranges.

I argue that the speed dating experiment fits these requirements perfectly. First, and most importantly, participants meet a sequence of partners and are asked to evaluate them after 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 results support the notion of relative thinking, as conceptualized by Bushong et al. (2015). The same person will hence be evaluated differently depending on the spread of the utility range, which in turn depends on previously considered participants. Evaluations translate into behavior (the decision to date) as follows: If the attribute is perceived positively, a larger range results in a *ceteris paribus* lower propensity to agree to a date, while the converse is true for levels of the attribute that provide disutility. This mirrors the mechanism Bushong et al. (2015) envision when claiming that fixed differences loom larger the more narrow the range against which they are evaluated.

On a more cautionary note, these results hold only for attractiveness, and seem to be much stronger in case the evaluator is female. They continue to hold qualitatively when considering the attribute *fun*, but are absent for the other three attributes on which participants of the experiments could rate each other. Reassuringly, neither of these latter attributes has any predictive power for the decision to date. As the identification strategy relies on picking up differences in the signs of the range coefficients across attribute levels, the absence of context dependence for these three attributes is expected.

To rule out other mechanisms that could explain the observed behavioral patterns, I control for contrast effects and time trends. In addition, I present a number of robustness checks, showing that the results are not driven by the specification of the attribute range I use throughout the main text, nor potential rescaling or calibration issues. None of these alternative specifications affect the main conclusion substantially. I therefore conclude that context influences decision makers, and behavior is in line with the notion of relative thinking proposed by Bushong et al. (2015). Considering that finding a suitable partner is a decision that should induce participants to put some thought into their choices and act according to their sincere preference, it is likely that context dependence also plays a role in situations with less serious consequences.

The three chapters of this dissertation are self-contained and can be read independently from each other. Appendices for all chapters are presented in Part iii, and a consolidated bibliography appears in Part iv.

Part II

MORALS, MARKETS, AND MALLEABILITY: THREE ESSAYS IN BEHAVIORAL ECONOMICS

1

ANIMAL WELFARE AND HUMAN ETHICS

1.1 INTRODUCTION

With technological progress, animal farming has changed considerably. Beginning in the late nineteenth century and paralleling the radical changes of the Industrial Revolution, intensive animal farming soon replaced more traditional forms of agriculture. Two discoveries were essential for this development: synthetic vitamin production enabled farmers to raise chickens indoors (Gordon, 1996), and animals needed to be given less space since contagious diseases could be contained by means of antibiotics and vaccines.

Despite its obvious advantages, including cheaper and more efficient production, it has long been recognized that intensive animal farming has severe deficiencies when it comes to ecological consequences and results in hardship for the animals involved. Some farmers have therefore returned to traditional forms of agriculture, which have become known as *organic* in order to highlight the distinction to intensive farming practices. At the same time, demand for organically produced food has increased substantially and continues to grow (cf. G. D. Thompson, 1998).

What motivates consumers to purchase organic food? Some may genuinely care about animal welfare and intend to support species-appropriate and sustainable farming. Yet increased demand for organic food could also be driven by other, self-oriented, reasons: customers may sometimes follow advertisements without much reflection. Furthermore, they may be motivated by health concerns, e.g. aiming at minimizing meat consumption of animals that received antibiotics. Consumers may also want to project a favorable image of themselves onto their peers if they are part of a subgroup that cares about animal welfare. In short, demand for organic products could result either from entirely selfish motives, environmental factors, or a genuine interest in animal welfare.[1]

To disentangle between these possible causes, it does not suffice to look at scanner data or aggregate market shares of organic products. These

1 We use the term 'animal welfare' to encompass a general disposition to care about living conditions of animals. The exact definition depends on our measure of the willingness to pay for better living conditions of animals and will be explained in detail below.

data are not informative about determinants of purchasing behavior. Likewise, hypothetical questionnaire studies have been criticized for measuring intent rather than behavior. In a rather stark example, Cowe and Williams (2000) find a preference for sustainable consumption among 30% of their sample in questionnaires, but only a 3% market share of goods that are produced according to ethical standards in the reference market (UK).[2] This suggests that questionnaire studies are misleading when it comes to measuring the market share of organic products. They may be even more inaccurate when assessing motivations for ethical purchase behavior, as social desirability would likely bias responses away from image concerns and towards more socially accepted motivations.

To close this gap in the literature and to better understand what motivates consumers to purchase organic foods, we elicit concern for animal welfare in an *incentivized, direct* and *real* setup, in a task that allows to abstract from self-oriented motives. Subjects choose between intensive farming and organic farming for a living laying hen. Opting for organic farming is costly, yet guarantees better food, daylight, and much more space to the hen. This procedure allows us to elicit subjects' willingness to pay for transferring a hen to an environment that is better suited to its needs.[3] Full anonymity and a single-blind experimental design ensure that our results are not diluted by image concerns or experimenter demand effects. We can therefore isolate concern for organic agriculture and animal welfare from other more self-oriented factors that could be at play when purchasing organic food at a supermarket.

2 This so-called "30:3 syndrome" or "ethical purchase gap" has been independently verified by other authors, such as Nicholls and Lee (2006). See Bray et al. (2010) for a review.

3 While opting for organic farming is a clear statement of opting for animal welfare, one may also see it as a decision in favor of sustainability. Defining sustainability is an elusive task. Yet one may consider a responsibility towards other species and the ecosystem as an important constituent of sustainable behavior. For instance, the UN follow a three pillar approach with regard to sustainability: "Consisting of three pillars, sustainable development seeks to achieve, in a balanced manner, economic development, social development and environmental protection." (http://www.un.org/en/ga/president/65/issues/sustdev.shtml). In that case, keeping animals in conditions in which they typically die from infections in very short time if they do not receive antibiotics in their feed, and in which they are deprived from nature and daylight, can easily be classified as non-sustainable. We acknowledge that it may not be possible to fully disentangle 'sustainability' and 'ethical responsibility'. These difficulties have been recognized and are widely discussed in the literature, for example by Singhapakdi, Vitell, Rao, and Kurtz (1999). For the purposes of this article, however, we remain agnostic about the distinction and view the decision about living conditions as a proxy for both.

The gap between questionnaire results and buying decisions may partly arise from the fact that it is cheaper to claim that one cares about animal welfare if it is costless (Bray et al., 2010; Carrigan & Attalla, 2001; Nicholls & Lee, 2006). This is why we establish a real tradeoff: money versus organic living conditions for a hen. Yet in addition to introducing money as an exchange medium, specific market structures may also affect decisions via other channels. Our setup deliberately abstracts from these channels, as their influence could easily vary depending on the specific market design.

For example, we abstract from the effects of advertisement. Through advertisement, people's attention may be directed to aspects of an animal product other than the living conditions of the animals. Similarly, firms may exploit consumers' limited availability of cognitive resources by making certain attributes of products more salient (Bordalo et al., 2013; Gennaioli & Shleifer, 2010). Furthermore, consumer behavior may be very different depending on whether consumers buy in a small shop or in a large supermarket. In the latter, they may feel less responsible for their buying behavior as they can share feelings of guilt with many other customers and receive more social information (Latané & Nida, 1981; Bandura, 1999; Rothenhäusler, Schweizer, & Szech, 2015). In addition, large markets extend the scope for consumers to apply replacement logics: Consumers might be prone to argue that even if they abstain from buying factory-farmed animal products, other consumers will buy. Similar reasoning leads to 'justification from substitution' (cf. Sobel, 2010; Falk & Szech, 2014).

We find that subjects are willing to pay 14 Euros for animal welfare on average. There is substantial heterogeneity in our measure: while 37% of subjects are willing to pay 25 Euros (the highest amount we allowed for) in order to put a laying hen into better living conditions, 15% are not willing to pay anything.

We also investigate what kind of people care about animal welfare when directly exposed to the decision between self-interest (i.e., money) versus improving the living conditions of a laying hen.

Intensive farming environments keep animals away from daylight and soil and provide only limited living space. Compared to traditional animal husbandry, intensive farming requires less human labor and allows farmers to produce more meat and other animal products in shorter time spans and in greater quantity. At first glance, intensive farming therefore seems to be an innovative and efficient way to structure production of animal products. It has even been argued that intensive farming may increase social welfare, as cheaper access to animal products could

improve human life (McCarthy & Bennett, 1986). Yet one could ask whether this reading of the term 'welfare' is broad enough: first of all, the ecological consequences of industrialized meat production seem worrying (Eisler et al., 2014). Emissions of greenhouse gases, for instance, would likely be substantially lower if animals were held in more appropriate environments, given that the supply of farming land is limited. The same holds for water use. Second, it is at least questionable whether eating animal products improves human health or not, specifically when they make up a large part of people's diets.[4]

Rather than contributing to the discussion about ecological and health-related consequences, we aim to understand the *social* importance of intensive farming. Put boldly: does the support of intensive farming go hand in hand with lower *moral* standards towards humans?

This question has received considerable attention from philosophers and ethicists, but clean empirical evidence is lacking. In a broad sense, whether there is an association between caring about animal welfare and high moral standards towards humans touches the philosophical debate asking where to legitimately draw ethical boundaries, i.e. which species to include in ethical considerations. Although our study does not claim to provide an answer to this very general question, we provide evidence that in the specific context we consider, enabling a laying hen to live in a more suitable environment goes hand in hand with a more empathic, prosocial, and value-oriented personality.

Philosophers like Aquinas, Locke and Kant hypothesized that there could be a link between the willingness to accept the suffering of animals and the willingness to accept human suffering. For example, Immanuel Kant argues that an appropriate treatment of animals is not a means in and of itself, but rather an instrument to ensure peaceful relations among *humans* (Adams et al., 2011). Aquinas postulates: "If in Holy Scripture there are found some injunctions forbidding the infliction of some cruelty toward brute animals [...] this is either for removing a man's mind from exercising cruelty towards other men [...] or because the injury inflicted on animals turns to a temporal loss for some man [...]." (Francione, 1995, p. 6). With regard to educating children, Locke advises parents that "[...] the custom of killing and tormenting of beasts, will, by degrees, harden their minds even towards men." (Locke, 1996, p. 90). More recently, Singer (1995) has argued that human ethics should extend to certain animals on the grounds that great apes, for instance,

4 For instance, a large, prospective study by Kelemen, Kushi, Jacobs, and Cerhan (2005) finds that deaths resulting from coronary heart disease decrease by 30% when study participants substitute vegetable for animal protein.

clearly show signs of intelligence and are similar to humans in their ability to suffer. Researchers have hence speculated that there could be a strong link between caring about animal welfare and human ethical rights.

If there is a link between treating animals well and respecting moral standards towards other humans, debate of intensive farming must include potential effects on human social behaviors, besides pure efficiency arguments. In fact, the notion of efficiency may be too narrow in this context, as moral consequences are outside its scope. However, the difference between humans and animals is often stressed as *the* demarcation line of ethics, suggesting that caring about animal welfare and behaving ethically towards other humans are two completely distinct issues.

For example, René Descartes proposed a mechanistic view of animals, which became rather popular in his time: To him, animals were automata, without soul or mind (Descartes, 1996). In many religions, animals are depicted as soulless creatures with few or no signs of intelligence. Aristotle established a clear hierarchy of humans over animals, arguing that the latter were made for the sake of human beings.[5] In addition, several defining aspects of humans, such as memory, self-awareness, and the ability to deliberate and carry out plans that reach far into the future, are frequently pointed out as constituting the boundary between beings that should and beings that should not be included in ethical considerations. Following such arguments, it has been concluded that ethical considerations should only apply to humans, and that the treatment of animals may be rather unrelated to how humans treat other humans (Wise, 2014). Our findings contradict the latter views in the following sense. There is a significant, positive association between caring about animal welfare and moral standards.

The setup of our experiment is simple. Participants are directly exposed to the decision between self-interest (money) versus improving the living conditions of a laying hen. If subjects opt for money, the hen will live under minimal legal standards, i.e., in intensive farming. If subjects forgo the money, the hen will instead live in an organic farming environment for the rest of its life, with access to daylight and soil, much more space, and organic feed without antibiotics.[6] Subjects know that

5 In contrast, Pythagoreans believed that souls could migrate from human to animal bodies. They accordingly stressed the importance of treating animals well (Wise, 2014).

6 The preventive use of antibiotics is ruled out by Naturland regulations. The regulations do allow for antibiotic treatment if animals become ill, but conditions are very strict: farmers must first use natural remedies. Only if these treatments do not take effect can antibiotics be used in very limited quantities. If laying hens are treated with antibiotic

the organic living conditions are certified and controlled by Naturland, a well-known labeling scheme for high organic farming standards in Germany.[7][8] We elicit decisions for various amounts of money in a price list, with monetary amounts increasing from 0.5 to 25 Euros. For each monetary amount, subjects decide between receiving money and exposing the hen to intensive farming conditions versus receiving no money and organic living conditions for the hen. Subjects know that one of their decisions is randomly drawn and implemented with all consequences.

As a measure of moral dispositions towards humans, we elicit subjects' tendencies to behave in Machiavellian ways. Machiavellianism, named after the 15th century philosopher and politician Niccolò Machiavelli, includes characteristics such as the willingness to manipulate others, behave opportunistically, neglect morals and show little affect in social interactions with other humans (Christie, Geis, & Berger, 1970). Machiavellianism has been measured in various studies and was found to correlate with economic opportunism (Sakalaki, Richardson, & Thépaut, 2007), lying (DePaulo & Rosenthal, 1979) and delinquency (Muris, Meesters, & Timmermans, 2013). Machiavellianism therefore serves as a validated measure of immoral and antisocial dispositions.

To enable a closer look at the personalities of our participants, we elicit neuroticism, extraversion, openness to experience, agreeableness and conscientiousness as part of the Big Five Personality Inventory. Agreeableness can be linked to compassion, empathy, altruism and an interest in other humans' well-being (Costa & McCrae, 1992). Agreeableness, like openness, is also associated with personal values such as humaneness and goodness (Olver & Mooradian, 2003), as well as prosocial motivations (Graziano, Habashi, Sheese, & Tobin, 2007). Openness has also been associated with empathy and emotional intelligence (Del Barrio, Aluja, & García, 2004; Van der Zee, Thijs, & Schakel, 2002). People who can easily imagine what life in intensive farming must be like for an animal may tend to save the hen from such living conditions. As such, more open personalities can be expected to have a higher willingness to pay for animal welfare. Thus, the openness-agreeableness nexus

substances more than once a year, their eggs cannot be labeled and sold as organic. (http://www.naturland.de/verbraucher.html).

7 "Naturland's farmers and processors work to the highest organic standards, which are even more stringent than the legal requirements imposed by the EC Organic Directive."(http://www.naturland.de/ourdistinguishingfeatures.html).

8 Of course, it remains debatable whether keeping animals in husbandry is acceptable at all. For a discussion, see, e.g., http://www.albertschweitzerfoundation.org.

serves as a second validated measure of other-regarding moral values and attitudes.[9]

Our data reveal that there is indeed a strong link between caring about animal welfare and standards in ethical behavior towards humans. Subjects willing to forgo higher amounts of money for the hen's sake score lower in the Machiavelli test and hence exhibit more moral, social dispositions towards humans. Furthermore, subjects who care more about the hen also show higher levels of openness in the Big Five Personality Inventory (Costa & McCrae, 1992). Hence, we find that people interested in animal welfare also tend to care more about the well-being of other humans and are more likely to have an altruistic, cooperative personality. In addition, we find that people oriented towards more market-friendly political parties care less about animal welfare and tend to opt for the selfish monetary amount. Maybe these subjects are used to thinking mostly about direct costs and benefits, leading them to ignore the well-being of a weak third party.

Our study fits into a recent literature discussing morality in economic transactions. Bartling, Weber, and Yao (2015) conduct two experiments in Switzerland and China to study whether participants have preferences for socially responsible behavior when interacting in markets. Their findings indicate that both consumers and producers are indeed willing to reduce negative externalities imposed on third parties, as in 45% of all transactions the more expensive but externality-free product is traded. Yet they also attest that the willingness to reduce externalities is higher still in non-market contexts, which speaks to the erosion of moral values through markets that had previously been confirmed by Falk and Szech (2013). In their experiment, subjects decide between saving the life of a mouse versus receiving 10 Euros and agreeing to kill the mouse. This task is used as a direct, incentivized measure of *moral transgression* (see Deckers, Armin, Kosse, and Szech (2016) for validation). While the task bears some resemblance to ours, killing a vertebrate without justification and appropriate qualifications is forbidden by law in many countries, including Germany.[10] In Falk and Szech (2013), subjects are

9 Olver and Mooradian (2003) concisely summarize research on these personality traits as follows: "[…] the extant empirical studies generally suggest that the more intellective traits of Openness, Agreeableness, and Conscientiousness relate systematically to Values. Specifically, trait Openness to Experience predisposes individuals toward values related to Openness to Change and Self-Transcendence (most specifically, Stimulation, Self-Direction, and Universalism) versus those related to Conservation and Self-Enhancement (Tradition, Conformity, Security, and Power)." (pp. 114).

10 Compare the German "Tierschutzgesetz" (Protection of Animals Act) as well as Art. 20a of the German constitution.

thus confronted with a decision that is perceived as morally relevant by many, that risks breaking some taboo in the sense that an animal is killed without intending to eat it or defending oneself against a potential threat, and that creates a strong sense of internal conflict. Deciding between money and better living conditions for a laying hen, in contrast, is very similar to many consumption decisions we face on a daily basis and does not violate any conventional norm. The two tasks therefore address different aspects of moral behavior.

1.2 EXPERIMENTAL DESIGN & PROCEDURES

1.2.1 *Methodology*

We use a controlled laboratory experiment to study whether subjects have a positive willingness to pay for animal welfare. After eliciting their willingness to pay, we ask participants to respond to well-validated and frequently used questionnaires in order to measure their ethical attitude towards other human beings. To the best of our knowledge, we are the first to use a fully incentivized, direct and real setup to study questions relating to animal welfare.

Previous studies on the relationship between personality traits, demographic factors and ethically responsible behavior are typically affected by some methodological features that are worth discussing in more detail. Most importantly, the dependent variable in the bulk of previous research has been constructed with the help of scenarios or vignettes.[11] As noted, for instance, by Marshall and Dewe (1997), the validity of the conclusions drawn from these types of hypothetical situations rests on two assumptions: (i), the scenario presented to respondents actually constitutes an ethical dilemma, and (ii), the context of this scenario is viewed in comparable ways by the respondents. In addition, the elicitation of moral behavior is not immune to social desirability concerns, rendering separation of intentions from actual behavior virtually impossible. In the context of ethical decision-making, this bias is especially worrisome. In hypothetical scenario and survey studies, the cost of presenting oneself as an individual that adheres to social norms is low. However, as evidenced by the gap between the willingness to engage in ethically

11 O'Fallon and Butterfield (2005) find that 55% of the 174 studies they analyze use hypothetical scenarios or vignettes in order to elicit a measure for ethically responsible behavior, while less than 4% of the results are based on lab studies or field experiments. Other ways to measure ethical behavior include asking subjects to assess their own behavior or the conduct of firms, as judged by internal audits (p. 404).

responsible purchase behavior and actual purchases, desired behavior does not necessarily match observed behavior. We address these potential concerns by designing an appropriate experimental environment and by introducing a trade-off between ethical behavior and monetary disutility.

1.2.2 *Design*

In order to elicit the degree to which subjects care about animal welfare, they trade off a monetary benefit against the welfare of a hen. Subjects decide between organic living conditions for the laying hen and forgoing a monetary amount versus conventional agriculture (intensive farming) for the hen and receiving a monetary amount.

Subjects know that if they opt for organic farming, the hen will live in a Naturland-certified farming environment. To be certified as organic egg producers, farms have to guarantee high standards with regard to ecologically sustainable production and are not allowed to add antibiotic substances to the feed. Subjects are informed that our farm is certified as an organic "Naturland" producer. This label is only given to farms fulfilling criteria that are substantially stricter than those prescribed by EU regulations. Conversely, conventional agriculture refers to legal minimum standards for livestock breeding.

Subjects know that the hen has been entrusted to their care. To make the decision more salient and to bridge the gap from the laboratory to a more natural environment, they are shown two pictures of a hen (Figure A.1 in Appendix A.1).

Moreover, the two possible living conditions, conventional versus organic farming, are described in detail and summarized in a table (Table 1.1).

After being exposed to the information on living conditions and the implications of their choices, subjects decide between monetary amounts versus animal welfare. They decide on a price list. In each decision row, they can opt for conventional living conditions for the hen and a monetary amount versus organic farming and no monetary amount. Monetary amounts increase over decision rows, from 0.5 Euros to 25 Euros in 0.5 Euro increments. Subjects know that one of their choices will later be randomly drawn in order to determine their payoff. They can hence guarantee organic farming to the hen by always choosing

	Conventional	Organic
grassy outdoor runs	no	~ 4 m^2 per hen
canopied outdoor area	no	available year-round, possibility to sand- and dustbathe
size of coop	up to 6000 hens	up to 3000 hens
feed	antibiotics allowed	no antibiotics allowed

TABLE 1.1.—Comparison of Living Conditions

the organic option. We include example payoff calculations and control questions to make sure that subjects understand the task.[12]

Choosing the price list method rather than a fixed monetary amount enables us to study subjects' willingness to pay for sustainable farming and a better life for the hen. The switch point, i.e. the amount of money at which subjects switch from organic to conventional agriculture, can hence be interpreted as how much money subjects are willing to forgo in order to enable a hen to live in a more sustainable and more appropriate environment.[13]

Since the consequences of their actions are not directly observable to participants in our experiment, we take care to state that their choices will be carried out exactly as specified in the instructions. To emphasize this point, we repeat the corresponding paragraph in oral form and inform subjects that they will be able to verify all facts after the experiment. To reduce experimenter effects, all sessions are conducted by a research assistant who is blind to the hypotheses of this study.

Another aspect of our study is worth noting. While subjects take an incentivized choice about the animal, we elicit moral dispositions towards humans via well-established test questions. We chose this combination of elicitation methods for the following reasons. If subjects first decided about forgoing money to improve an animal's welfare and then took a real-life decision in an ethical context (e.g. forgoing money to help a human being versus taking the money), the first decision could easily affect the second. This so-called *conscience accounting* is reminiscent

12 Translated instructions are included in Appendix A.1. Instructions in the original language (German) are available upon request.

13 Subjects who always choose the monetary reward (i.e., never choose organic farming) are assigned a switch point of 0.5 Euros. Accordingly, subjects who always choose organic farming are assigned a switch point of 25.5 Euros.

of the trade in pardons used by the medieval Catholic church to raise money, and has recently been confirmed experimentally by U. Gneezy, Imas, and Madarász (2014). In particular, if subjects made a strong link between animal welfare and human ethics, which is what we aim to explore in this study, subjects with a bad conscience from harming the animal could decide to help the human in the second step just to ease their bad conscience and thereby obtain a better self-image (Bénabou & Tirole, 2011). This is why we rely on validated questionnaires to elicit moral and prosocial dispositions instead of providing a second task with another real third party that may be helped or harmed.

1.2.3 *Procedures*

The experiment was conducted at the University of Bamberg on two consecutive days in February 2013 with a sample comprising predominantly students. In total, 216 subjects (117 female, 94 male, 5 unidentified) participated in one of five sessions with an average duration of 105 minutes. Average earnings were approximately 13 Euros, including a show-up fee of 4 Euros. Participants were recruited via ORSEE (Greiner, 2004).

Subjects received written instructions for the experiment. Before receiving the instructions, they read a text on free will, religion, or brain activities. Texts were randomly allocated and constitute our treatment conditions. Since switch points do not differ across conditions, we collapse the data in order to obtain a broader basis for our study (see Appendix A.2 for details on the text treatments and statistical tests showing that p-values are not significant on any conventional level). All analyses are replicated on the control sample (see Appendix A.3).

After all subjects had filled out the price list, they were asked to complete the Big Five Personality Inventory (Costa & McCrae, 1992), a test for Machiavellian personality traits (Christie et al., 1970), the Bem Sex Role Inventory (Bem, 1981) and a series of demographic questions. In order to motivate subjects to respond to these questions with proper attention, we incentivized this stage by paying an additional 3 Euros.

1.2.4 *Implementation*

In order to implement the decisions taken by the subjects in our experiment, we cooperated with two local egg farmers. One of these egg producers keeps the laying hens in large coops that are run according to conventional standards. The other, neighboring, egg farmer is certified

as organic according to the high standards of Naturland. We arranged that for each hen involved in our study, two places were kept open: one in intensive farming, one in the organic environment. We opted for this solution as it ensures that there is no scope for a replacement argument, meaning that the remaining capacity is not filled up with other hens. For each hen, there are two living options, and the one that is not chosen by subjects remains unfilled.

1.3 PREDICTIONS

Our research design allows us to study correlations between individual characteristics and our measure of ethical behavior, the switch point. We derive hypotheses concerning the direction of effects in the following section.

1.3.1 *Personality Traits*

Personality traits are generally more difficult to observe than demographic and socioeconomic information, yet they often exhibit a higher degree of predictive power (De Pelsmacker, Driesen, & Rayp, 2005; O'Fallon & Butterfield, 2005). The availability of validated and often used scales enables us to identify key personality factors that are likely to play a role for moral judgment. We concentrate on two standard measures: Machiavellianism (Mach-IV) and the Big Five Personality Inventory (NEO-FFI). Both have been extensively analyzed in various contexts. We also include the Bem Sex Role Inventory (BSRI) to explore whether self-ascribed gender roles have an impact on willingness to pay for animal welfare, over and above that of biological gender.

Machiavellian characters exhibit a high willingness to manipulate others, behave opportunistically, neglect morals and show little affect in social interactions (Christie et al., 1970). Machiavellianism as a personality construct is generally considered to consist of three dimensions, reflecting different themes in the original writings upon which the Machiavelli scale is based: "The first theme was the endorsement of such manipulative tactics as the use of flattery and deceit in interpersonal interactions. The second theme was a cynical view of human nature in which others are regarded as weak, untrustworthy and self-serving. The third theme was a disregard for conventional morality." (B. Fehr, Samsom, & Paulhus, 2013, p. 78). The Mach-IV scale we use accounts

predominantly for the first two themes.[14] We therefore compute scores for 'tactics' and 'cynicism' separately and include them as two regressors in our analyses.[15]

Machiavellianism has been shown to be negatively associated with ethical decision making and behavior (Ford & Richardson, 1994; O'Fallon & Butterfield, 2005). People who score high on the Machiavellianism scale are more likely to evaluate unethical behavior as appropriate if it is effective in achieving a personal goal than people with low scores (Geis & Moon, 1981; Singhapakdi & Vitell, 1990). Furthermore, machiavellistic people exhibit a higher propensity to lie (DePaulo & Rosenthal, 1979). Kahane et al. (2012) suggest that such unethical behavior may be rooted in a lack of empathy and lack of aversion to harm others. Thus, if there is a link between moral standards in social interactions with humans and caring about animal welfare and sustainability, Machiavellianism should be negatively associated with decision making that enhances animal welfare and provides more sustainable living conditions to animals. We accordingly expect subjects with high Machiavellianism scores to care less about the living conditions of their hen.

Of the five personality dimensions that Costa and McCrae (1992) identify as basic traits, openness and agreeableness are most closely related to empathic behavior and prosocial motivation. In their Big Five inventory, Costa and McCrae (1992) describe an agreeable person as being sympathetic, interested in the well-being of others and compassionate. They further link agreeableness to characteristics such as empathy and altruism. It is hence not surprising that agreeable individuals are reported to be prosocially motivated (Graziano et al., 2007; Wilkowski, Robinson, & Meier, 2006).

Openness as a concept is rather difficult to grasp, but there is wide-ranging agreement that it refers to a personality with strong beliefs in values, a perceptive and curious intellect, and a liberal, adventurous, and empathic mindset (McCrae & Costa, 1997). The link between openness and empathy has also been established by Del Barrio et al. (2004), and it has been shown that subjects who score high on the openness facet possess higher levels of emotional intelligence (Van der Zee et al., 2002).

Taken together, these results suggest that if there is a relation between moral behavior and caring about animal welfare, subjects scoring high on

14 Only two out of 20 questions refer to the morality dimension. Any measurement of this subfactor by means of the Mach-IV scale is therefore bound to suffer from low precision.

15 This procedure is also recommended by B. Fehr et al. (2013), who provide a thorough review of the literature spawned by the Machiavellianism scale from 1971 to 1987.

agreeableness and openness will make more animal welfare-enhancing and sustainable decisions. We thus hypothesize that agreeableness and openness will correlate positively with the amount of money subjects are willing to forgo to increase a hen's living conditions.

An exploratory questionnaire we use in our study is the Bem Sex Role Inventory (BSRI). The BSRI is a very useful instrument to abstract from biological gender and instead focus on self-ascribed sex roles that range from rather masculine to rather feminine. The questionnaire consists of a series of attributes that are usually associated with feminine or masculine behavior. Several questionnaire studies have suggested that females may behave in more ethical ways than males, while other studies find no significant correlation between biological gender and morality (for instance, Chung and Trivedi (2003) and Ross and Robertson (2003) report significant results confirming the positive association between female gender and morality, while in the majority of scenarios in Lund (2000) and Radtke (2000) there is no gender difference in behavior). Besides biological gender, we are interested in sex roles in order to see whether self-ascribed gender may play a role for decision-making.

1.3.2 *Demographic Variables*

As indicated above, demographic criteria are often found to possess low explanatory power for ethically responsible or sustainable behavior. For instance, a study by Straughan and Roberts (1999) examines the link between demographic and psychographic criteria with ecologically conscious behavior. They find that psychographic criteria are much better suited to explain observed variation in the dependent measure than demographic factors.

Our results are largely in line with this general tendency. Across the spectrum of demographic questions we posed, only those that are highly correlated with personality traits (like being a vegetarian) possess predictive power.

1.4 RESULTS

We define the switch point to be the first decision where a subject switches from the organic to the conventional agriculture option. A total of 16 out of 216 subjects switch between the two options multiple times, and are classified accordingly. We exclude 19 out of the remaining 200 subjects because they switch from conventional to organic farming.

	MEN			WOMEN		
	∅	σ	N	∅	σ	N
Age	23.79	3.09	84	23.38	3.50	106
Vegetarian	0.10	0.30	84	0.20	0.40	106
Vegan	0.01	0.11	84	0.01	0.10	106
Religiosity Index	0.39	0.20	84	0.38	0.19	106

TABLE 1.2.—Summary Statistics

These subjects either preferred intensive farming to organic farming, or they did not understand the task. Including multiple switchers, this leaves 197 subjects in the sample. If subjects switch more than once, we impute their willingness to pay by taking the first switch point as a proxy. As a robustness check, we also present analyses on the sample without multiple switchers, which comprises 181 subjects.

Table 1.2 summarizes main demographic characteristics of the subjects used in this study.

In order to test our predictions, we present regressions with the willingness to pay for animal welfare as dependent variable and personality measures as well as demographic information as regressors. As is evident from Figure 1.1, the dependent variable is censored both from above and from below.[16] This would result in an inconsistent estimator when using OLS. To deal with censored data, Tobin (1958) proposed a maximum likelihood estimator that Amemiya (1973) subsequently proved to be consistent. We therefore report regressions using the Tobit model.

1.4.1 *Willingness to Pay for Animal Welfare*

As pointed out before, the nature of our task allows for a clean identification of willingness to pay for animal welfare. We explicitly abstract from advertisement and replacement effects and suppress image concerns by ensuring full anonymity. Our data indicate that the median subject is willing to pay 11 Euros for animal welfare (the average WTP is 14 Euros).

16 The empirical probability density function of the switch point has mass points both at 0.5 Euros and at 25.5 Euros. It is therefore likely that a significant fraction of subjects would have been willing to pay more than 25.5 Euros in order to ensure better living conditions for the hen, while an equally significant (yet not quite as large) fraction of subjects may exhibit a negative willingness to pay for animal welfare.

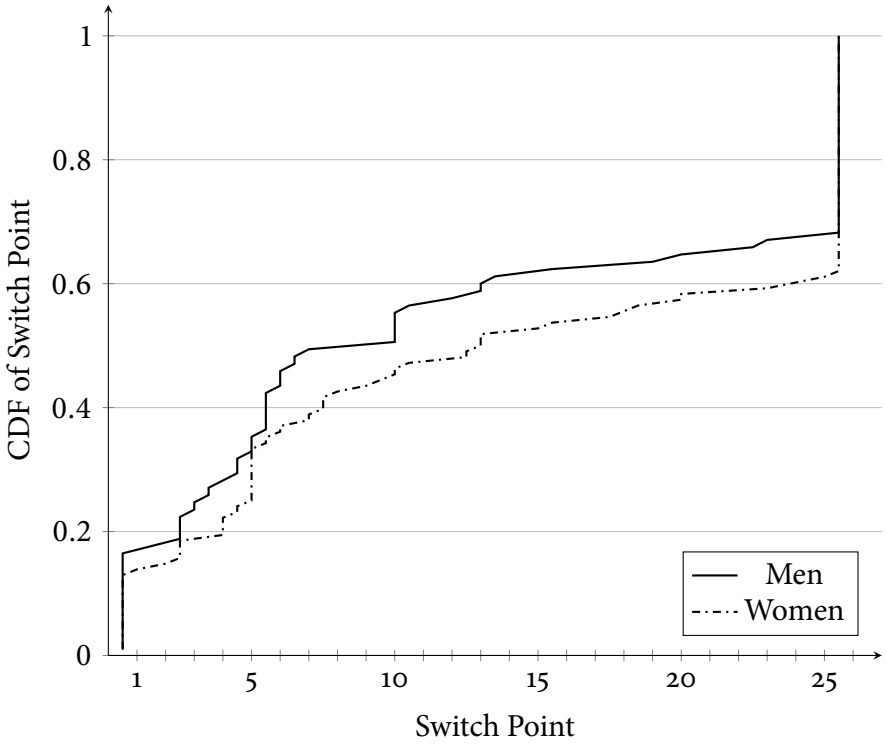


FIGURE 1.1.—Willingness to Pay for Animal Welfare

Notes: Empirical cumulative density function of the switch point, separately by gender.

Visual inspection of Figure 1.1 reveals substantial heterogeneity in the switch point. About 37% of participants refuse to accept conventional farming for any amount of money we offered, thereby ensuring organic farming for their hen. On the other end of the spectrum, 15% of subjects are unwilling to give up any amount of money in order to ensure better living conditions for the laying hen. There is no statistically significant difference between men and women regarding the median and the distribution of the switch point.

As discussed in Section 1.1, a number of studies have shown a gap between intent to behave in ethical ways and actual behavior in the marketplace (cf. Cowe & Williams, 2000; Nicholls & Lee, 2006; Bray et al., 2010). We cannot directly attest to this 'ethical purchase gap', since we lack data on intent. However, according to personal communication with the two cooperating farmers, the average payment of 14 Euros corresponds almost exactly to the additional cost imposed by organic farming relative to conventional agriculture. This suggests that participants have a realistic understanding of the two different conditions and are willing to compensate farmers for additionally incurred costs in full. We may therefore speculate that in our sample, the willingness to pay for organic farming conditions is sufficiently high in order to induce the majority of subjects to buy organic even in the marketplace.

1.4.2 *Animal Welfare and Human Ethics*

The central question we aim to answer with our study concerns the social importance of intensive animal farming. Our results show that lower moral standards towards humans are indeed associated with less concern for animal welfare. This enables us to speak to a long-running debate about the relationship between animal and human cruelty. For instance, Kant makes the following *empirical* claim: "If he is not to stifle his human feelings, he must practice kindness towards animals, for he who is cruel to animals becomes hard also in his dealings with men." (Gruen, 2014). It is certainly true that Kant claims a causal relationship that we cannot show. However, we do fail to reject the hypothesis by showing that there is a correlation between animal welfare and ethical standards towards humans.

Put differently, we provide empirical support for Kant's claim that "we can judge the heart of a man by his treatment of animals" (Kant, 1981), which is a necesary condition for the claim that cruelty to animals leads

	Excl. Multiple Switchers		Incl. Multiple Switchers	
	(1)	(2)	(3)	(4)
Openness	0.549**	0.473*	0.517**	0.422*
	(0.250)	(0.259)	(0.218)	(0.230)
Agreeableness	0.431	0.295	0.246	0.178
	(0.416)	(0.414)	(0.344)	(0.338)
Machiavelli (Tactics)	-0.801**	-0.707*	-0.680**	-0.548
	(0.383)	(0.401)	(0.320)	(0.336)
Machiavelli (Cynicism)	0.557	0.458	0.302	0.163
	(0.486)	(0.484)	(0.388)	(0.386)
Locus of Control	-0.304	-0.223	-0.286	-0.210
	(0.230)	(0.227)	(0.202)	(0.199)
Masculinism Score	0.158	0.0512	0.143	0.0559
	(0.192)	(0.188)	(0.174)	(0.171)
Feminism Score	0.217	0.203	0.163	0.155
	(0.192)	(0.200)	(0.166)	(0.169)
Age		0.580		0.762*
		(0.536)		(0.459)
Gender		-3.363		-2.590
		(3.733)		(3.241)
Vegetarian or Vegan		12.47**		9.178*
		(5.812)		(4.700)
Religiosity Dummy		-1.150		-1.352
		(3.600)		(3.021)
Constant	-13.86	-16.85	-0.00792	-10.50
	(31.32)	(31.07)	(26.42)	(26.28)
N	153	147	168	162
N (uncensored)	68	66	83	81
N (left-censored)	24	22	24	22
N (right-censored)	61	59	61	59
Pseudo R^2	0.021	0.034	0.019	0.029
Log Likelihood	-370.8	-353.4	-432.1	-414.7

TABLE 1.3.—Determinants of Concern for Animal Welfare

Notes: Tobit regressions with switch point as dependent variable. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

to cruelty towards humans, and an important link in his argument as to why humans should be good to animals.

As measures for ethical dispositions towards humans, we elicited two dimensions of Machiavellianism: a cynical view of human nature and a general disposition to exploit human weakness for personal gains through manipulative tactics (B. Fehr et al., 2013). The first dimension concerns the perception of others without appealing to the moral views of the respondent himself. The second dimension prods subjects for their willingness to participate in exploitative endeavors, essentially asking them for their willingness to circumvent conventional moral standards to achieve personal gains. We therefore focus on the second dimension (which we term Machiavellianism (Tactics) in the regressions below) to gain insights into subjects' personal attitudes regarding ethical behavior. In addition, we asked subjects to complete the full Big Five personality inventory and use their agreeableness and openness scores as proxies for a general disposition to respect ethical standards and endorse humane and good behavior (Olver & Mooradian, 2003).

Table 1.3 presents our results. Exactly as predicted, we find a significant negative correlation between the switch point and the tactics dimension of Machiavellianism: The more money a subject is prepared to forgo in favor of organic farming, the higher are his or her standards concerning morals and other-regarding behavior in social interactions with humans. As hypothesized, people who classify as detached from conventional morality and unemotional according to the Machiavellianism (Tactics) scale care less about the hen's living conditions than other subjects. The effect is statistically and economically significant: if the Machiavelli (Tactics) score increases by one point, willingness to pay for animal welfare decreases by 0.8 Euros on average. Measured against the empirical range of scores (with a minimum of 11 and a maximum of 43), this implies that a 4 percentage point increase in Machiavellianism results in a 1 Euro decrease in the switch point. To illustrate the magnitude of this effect, someone in the first quartile of the distribution would be willing to pay at least 5 Euros more than a person in the fourth quartile.

As a second measure of moral dispositions towards humans, we elicited the subject's scores in the Big Five facets agreeableness and openness. We expected a strong link between agreeableness and caring about animal welfare. Our data do not support this hypothesis. This may be due to several factors: either the measure of agreeableness as personality construct is imprecise, or it does not measure what it claims to measure, or agreeableness is indeed uncorrelated with caring about animal welfare. Considering the relatively small sample size, we may

also not have enough power to detect the effect if it is sufficiently small. The highly significant negative correlation ($\rho = -0.39$, $p < 0.01$) between the Machiavellianism (Tactics) score and Big Five facet agreeableness lends some support to the latter explanation.

Reassuringly, we do find a link between openness and concerns for animal welfare: a two point increase in openness is associated with a 1 Euro increase in willingness to pay. In terms of economic significance, this effect is very similar to the Machiavellianism effect: subjects in the third quartile of openness are willing to pay at least 5 Euros more for better living conditions than those in the first quartile. Openness is not only associated with characteristics such as creativity, attention to inner feelings and tolerance, but also with goodness and humaneness (Olver & Mooradian, 2003), which are likely to affect behavior towards others. Further, openness is associated with cognitive abilities (DeYoung, Peterson, & Higgins, 2005). Previous research has reported correlations between cognitive ability and the ability to delay gratification (Dohmen, Falk, Huffman, & Sunde, 2010), i.e., the ability for future oriented thinking. Cognitive ability, and hence openness, might accordingly also be connected to ecologically sustainable behavior, for which considerations of future outcomes play an important role.

Our results are in line with findings from psychology and sociology that focus on whether there is a link between aggressive behavior towards animals and aggressive behavior towards humans. Research from psychology, psychiatry and sociology suggests that children's aggression towards animals predicts violence against human beings later in life (Felthous & Kellert, 1986; Flynn, 1999, 2000). In a review study, Gullone (2011) investigates the co-occurrence of human-directed and animal-directed aggression and concludes that children and adults who behave aggressively against animals are also likely to act aggressively and show violence against people. Yet if there is a link with regard to aggression, there could also be a link with regard to compassion. K. L. Thompson and Gullone (2003) argue that contact to animals and developing a bond with them enhances empathy towards animals in children, which will be transferred to human beings and thus lead to higher prosocial behavior towards humans. Our study shows that indeed, caring about animal welfare correlates strongly with moral dispositions towards human beings. There seems to be a strong relationship between human ethics and a preference for more sustainable, more appropriate living conditions for animals. Conversely, if people are prepared to look away when the welfare of an animal is concerned, they are also likely to look away when a human needs their support and cooperation.

1.4.3 *Further Exploratory Findings and Validity of Measure*

VEGETARIANISM As a validity check, we investigate the association of the switch point with being a vegetarian.[17] Following a vegetarian diet can be driven by different motives, including health concerns or religious convictions. Therefore, being a vegetarian does not necessarily have to coincide with caring about animal welfare. We nevertheless expected that many vegetarians would likely have a preference for improved living conditions for the hen. A Wilcoxon rank-sum test confirms that vegetarians show a higher willingness to forgo money for the hen's sake than non-vegetarians ($\text{mean}_{\text{non-vegetarian}} = 12.84$, $\text{mean}_{\text{vegetarian}} = 16.68$, $p < 0.1$). This association is also apparent in the regressions reported in Table 1.3. We take this as a sign of validity of our measure of preferences for animal welfare.

GENDER DIFFERENCES While the literature on moral behavior does not always reveal a gender difference, it is a standard finding that *if* there is a difference in gender with regard to ethical or moral behavior, females have the higher standards.[18] When it comes to caring about animal welfare as measured by switching from organic to conventional farming, we do not observe an overall significant difference between genders ($\text{mean}_{\text{female}} = 14.29$, $\text{mean}_{\text{male}} = 12.48$, $p = .26$).

Likewise, we fail to show that feminity (masculinity) assessed with the Bem Sex Role Inventory correlates positively (negatively) with the switch point (see Table 1.3). Thus, neither biological sex nor self-ascribed sex roles seem to be associated with the willingness to pay for sustainability and animal welfare.

POLITICAL ORIENTATION We do find that switch points differ depending on the political party subjects identify with. The parties covered the entire political spectrum, with the *Christian Democratic Union* (CDU) being the most conservative option, and the party *The Left* (Linke) on the far left. Subjects opting for the CDU are likely to have a traditional, conservative view on social life and a market-friendly attitude. The *Free Democratic Party* (FDP) is a traditional liberal, market-friendly party; *The Green Party* (Gruene) follows policies directed at en-

17 We did not further differentiate between pesco-vegetarians, ovo-lacto-vegetarians etc. in our questionnaire. We assessed veganism separately, but include these subjects in the vegetarian category.

18 Compare the survey articles by Loe, Ferrell, and Mansfield (2000) and O'Fallon and Butterfield (2005).

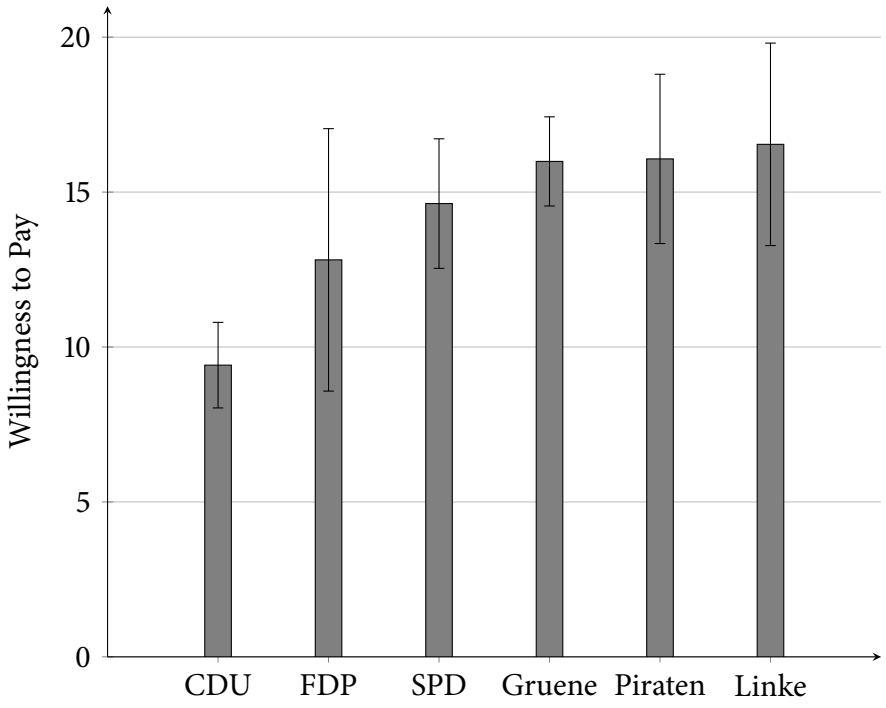


FIGURE 1.2.—Animal Welfare and Political Orientation

Notes: Average switch points by identification with political parties. Bars display standard error of the mean.

vironmental protection, while the *Social Democratic Party* (SPD) is the traditional socialist party in Germany. The party *The Left* is a left-wing party with a sceptic attitude towards capitalism, and the *Pirates* (Piraten) represent a more direct democracy and open internet culture.

Interestingly, political orientation bears some explanatory power about who cares about animal welfare. Our results suggest that a tendency towards market-friendly policies coincides with a low interest in animal welfare, while subjects with a more socialist or environment-oriented political view care more about the living standards of animals (Figure 1.2).

There is an overall significant difference between switch points according to political orientation (Kruskal-Wallis test, $p < .05$). Specifically, subjects with a preference for rather market-friendly parties (i.e. CDU and FDP) are less prepared to forgo money for organic living conditions compared to subjects preferring any other party.[19] The fact that conven-

19 CDU/FDP vs. SPD: $p < .05$; CDU/FDP vs. Gruene: $p < .01$; CDU/FDP vs. Piraten: $p < .05$; CDU/FDP vs. Linke: $p < .1$.

tional animal farming is the standard in Germany might serve as an indication as to why subjects with a preference for the conservative party CDU seem to care least about their hen's welfare. Most farm animals have been kept in factory farming conditions over the last decades in Germany.[20] Hence, following the conventional approach could imply a preference for keeping animals' living conditions *as is*.

1.5 CONCLUDING REMARKS

The perspective that animals do not possess innate rights to be treated well has a long tradition in philosophy, going back to Thomas Aquinas (1225-1274). Over the course of the centuries, philosophers like Kant and Locke have confirmed these views, placing animals firmly in the realm of creatures that deserve to be treated well only if it is in the interest of *humans*. Thus, until not too long ago, there has been a consensus that ethical considerations should include only other human beings and leave animals outside. By showing that caring about the conditions farm animals live in directly relates to acting in accordance with social and moral norms in human interaction, our results caution against this view.

Our aim in this study was to investigate whether there is a clear link between caring about animal welfare and having high ethical standards in general. Our data confirm that this relation exists. People who care about animal welfare also express higher ethical standards towards human beings.

The design of our study incentivizes decisions about animal welfare by establishing a trade-off between a monetary benefit and the possibility of enabling a laying hen to live in a more appropriate environment. In contrast to scenario or vignette studies that cannot avoid biases due to social desirability concerns, our design therefore allows for a clear identification of subjects' interest in animal welfare, abstracting from self-related aspects like image concerns or social desirability.

Several philosophical treatises have speculated that it is important to treat animals well because such behavior stimulates ethical conduct among humans. Our study stresses that such a link could exist. To our knowledge, this experiment is the first to isolate interest in animal welfare from confounding factors like the market environment and to establish

20 According to the German Statistical Office, only 16.7% of laying hens were kept in free range or organic farming conditions at the time of the latest agricultural census (https://www.destatis.de/DE/ZahlenFakten/Wirtschaftsbereiche/LandForstwirtschaftFischerei/Landwirtschaftszaehlung2010/Tabellen/9_3_LandwBetriebeHaltungsplaetzeHuehner.html).

a clean link with Machiavellianism and agreeableness, our measures for ethical behavior towards humans.

In addition, we conducted exploratory analyses, examining the relations between various individual characteristics and caring about animal welfare in order to inform policy makers and social debate. We find that there are pronounced differences in caring about animal welfare depending on which political party subjects like best. Subjects preferring more conservative and market-friendly parties care less about animal welfare. Furthermore, open-minded subjects and participants scoring low on the Machiavellianism (Tactics) scale care more about animal welfare than other subjects. Neither biological gender nor self-ascribed sex roles are good predictors for ethically responsible behavior.

Beyond offering a segmentation of customers by ethical responsibility, we do not comment extensively on measures a policy maker could take in order to induce people to care more about animal welfare. However, we think that this question is a fascinating topic for future research.

2

DELEGATING PRICING POWER TO CUSTOMERS

2.1 INTRODUCTION

In many service industries such as hotels, airlines, or entertainment, firms have large fixed costs for capacity but face widely fluctuating demand. Sophisticated methods of price discrimination such as yield management have been developed to use capacity optimally by setting (often vastly) different prices in periods of high and low demand (Weatherford & Bodily, 1992). However, offering low prices in reaction to off-peak demand bears the risk that consumers' reference prices are affected, which may significantly reduce their willingness-to-pay in periods with regular and peak demand (Kalyanaram & Winer, 1995).[1]

Customer-driven pricing mechanisms are a potential solution to this problem. By delegating the pricing decision to customers, they price discriminate without setting a reference point: sellers using *Pay What You Want* (PWYW) ask consumers to pay any price they like, including zero (Schmidt, Spann, & Zeithammer, 2015). Sellers using *Name Your Own Price* (NYOP) ask consumers to submit a bid against a threshold that is set by the seller but unknown to buyers. A transaction takes place only if the offered price exceeds this threshold (Spann & Tellis, 2006). NYOP was invented by Priceline, an online travel intermediary selling flights, hotel rooms, rental cars and vacation packages in the United States (Anderson, 2009; Dolan & Moon, 2000).[2] PWYW has been applied mainly in service industries such as hotels, restaurants or museums (Kim et al., 2009).[3] These pricing mechanisms achieve endogenous price discrimination because different customers pay different prices depending on their

1 For example, hotels in New York City experience low booking rates especially in January and February and apparently try to increase capacity utilization by lowering prices. See, e.g., http://www.nycedc.com/economic-data/travel-and-tourism. However, prices are not decreased enough to make full use of capacity, which may be explained by concerns about reference price effects.

2 Priceline's business model has been very successful, with a current market cap of $65.5 bn and gross profit of $7.6 bn (http://finance.yahoo.com/q/ks?s=PCLN+Key+Statistics). Other current examples of NYOP sellers include the Danish website prisminister.dk, or eBay sellers selling via eBay's Best Offer option.

3 See for example http://www.bbc.com/travel/feature/20140730-pay-what-you-want-at-a-paris-hotel,

valuations, their conception of fairness, their beliefs about acceptable bids, or their degree of risk aversion.

We argue that customer-driven pricing mechanisms offer direct and indirect promotional benefits to sellers. Their participatory and innovative nature appeals to many consumers and often initiates word-of-mouth recommendations and favorable press coverage (Hinz & Spann, 2008; Kim et al., 2009). Thus, a direct benefit is that they can be a powerful tool to promote a product to a wider audience, in particular if the firm has a strong social media position with many fans and followers.

Furthermore, PWYW and NYOP may carry indirect promotional benefits by increasing demand for complementary products. For instance, the rock band Radiohead pioneered the use of the PWYW format on the Internet, giving consumers the opportunity to pay what they saw fit for their album *In Rainbows*. Even though profit margins on the album itself may have been larger had they used posted prices, by choosing PWYW they were able to attract many new followers and increase demand for complementary products. According to Chesbrough (2010), "[w]hatever revenue Radiohead might have lost through its initial download experiment was more than compensated for by the far greater publicity the band received, which seems to have accounted for the surge in commercial sales, and no doubt also benefited ticket sales for its subsequent world tour" (p. 357).

PWYW and NYOP can also be used as a temporary promotional tool to attract customers. For example, the ibis hotel Chennai City Centre has offered Pay What You Want-Deals for a limited travel period. The deal was advertised as 'special offer' on the hotel's website, presumably to increase brand awareness and thereby traveller demand in non-PWYW periods.[4] Even though this may initially result in losses, they can be overcompensated if customers continue to frequent the service once it has reverted to regular posted prices. In contrast to the word-of-mouth effects discussed above, which target a previously untapped customer segment, here additional benefits arise from intertemporal spillover effects. Rather than catering to a different customer segment, intertemporal spillover effects therefore increase demand from the *same* customers at later points in time.

This effect need not be limited to PWYW: for a given threshold, a NYOP seller will attract weakly more demand than a posted-price (PP) seller with an equally high posted price if consumers underestimate the

http://www.freetoursbyfoot.com/new-york-tours/,
http://www.accorhotels.com/owm002740-001-pay-what-you-want.shtml.

4 See http://www.accorhotels.com/owm002740-001-pay-what-you-want.shtml.

threshold on average.[5] Even though not all of these additional consumers submit successful bids, this provides sellers with the opportunity to cross-sell alternative products.

In this chapter we compare PWYW and NYOP to each other and to traditional posted prices. We use controlled lab experiments to identify the driving forces of the behavior of buyers and sellers who are facing or employing these mechanisms. What determines how much buyers pay? Under what circumstances are PWYW and NYOP viable and when do sellers choose to use them? In particular, what role do direct and indirect promotional benefits play with regard to the profitability of such mechanisms? How can they be employed as a competitive strategy to capture market share and to reduce competition? Do word-of-mouth benefits arise endogenously?

We find strong and important differences between PWYW and NYOP. PWYW appeals to all customers and achieves (almost) full market penetration. It can be used as a very aggressive strategy that drives competing posted price sellers out of the market. Many consumers pay positive prices voluntarily, but PWYW is profitable only if costs are low and if there are additional benefits (e.g., press coverage, word-of-mouth recommendations, or spillover effects on complementary products). In contrast, NYOP is a much less aggressive strategy. It relaxes price competition and leaves a significant share of the market to the competing posted-price seller. It is particularly appealing to low valuation customers who do not lose much if their bid is unsuccessful and for whom the price offered by a posted-price seller is too high. High valuation customers, on the other hand, prefer to buy from the posted-price seller in order to avoid the risk that their bid is unsuccessful. NYOP can also be used for goods with high marginal costs and if there are no additional benefits because the seller can protect himself against selling his good below cost by setting an appropriate reserve price.

In a second set of experiments we endogenize direct promotional benefits due to media buzz and word-of-mouth advertising. We show that these benefits differ across PWYW and NYOP. For NYOP the additional benefits triggered by word of mouth are substantial. NYOP sellers are able to capture a larger market share and to make higher profits than traditional posted price sellers. In contrast, PWYW is even more successful in capturing market share and monopolizes the market, but the high market share does not translate into high profits because consumers

5 Even if consumers have correct beliefs about the threshold in relation to the posted price, it is almost costless to submit a bid at the NYOP seller before purchasing at posted prices.

are paying too little. We conclude that indirect promotional benefits are required for the viability of PWYW.

So far the literature on customer-driven pricing mechanisms has dealt with PWYW and NYOP separately. There are many case studies and field experiments on PWYW in specific industries. These include the music industry,[6] zoos and museums,[7] restaurants and wine bars,[8] online content,[9] hotels (Gautier & Klaauw, 2012) and travel agencies (León, Noguera, & Tena-Sánchez, 2012), souvenir photos in an amusement park (A. Gneezy, Gneezy, Nelson, & Brown, 2010), or the Google service Google Answers (Regner, 2014). These studies are highly instructive but they cannot identify the causal effects driving behavior in these markets.

There are a few laboratory studies that allow for more control of the environment. For example, Mak, Zwick, Rao, and Pattaratanakun (2015) use a laboratory experiment to study how subjects can coordinate their behavior to make PWYW viable even if they are not motivated by social preferences. Schmidt et al. (2015) also use laboratory experiments on PWYW markets to identify the causal effects that determine voluntary payments both in monopolistic and competitive markets. However, they focus on an environment in which (a few) buyers and a seller interact repeatedly, so that buyers have an incentive to keep the seller in business. This is an important concern for neighborhood restaurants and other local service industries with a stable customer base. In contrast, the present chapter is concerned with industries in which firms face an anonymous customer base with one-shot interaction. With one-shot interaction between buyers and sellers it is much more difficult for PWYW to be viable.

A key innovation of the present study is the introduction of additional benefits that accrue to sellers employing a customer-driven pricing

6 PWYW has been used by the bands "Radiohead", "Nine Inch Nails", and "Moby" (Johnson & Cui, 2013) as well as by the online music label "Magnatune" (Regner & Barria, 2009).

7 The zoo in Augsburg, Germany has successfully employed PWYW during the advent season in 2013, 2014 and 2015. One reason for doing so is to compete for consumers' attention when there are many other distractions (christmas markets etc.) on offer. This strategy seems to work well: according to the data we received, visitor numbers tripled in PWYW periods (relative to the same period in previous years). PWYW is also employed by many museums such as the Metropolitan Museum of Art in New York or the Museum König in Bonn, Germany.

8 Examples include the restaurant *Kish* in Frankfurt, Germany (Kim, Natter, & Spann, 2010), or *Der Wiener Deewan* in Vienna, Austria, and wine bars like *Weinerei* in Berlin, Germany.

9 Wikipedia and Humble Bundle are a case in point.

mechanism. We not only examine how additional benefits affect market outcomes when at least one seller uses PWYW or NYOP, but we also show that such benefits can arise endogenously. Finally, to the best of our knowledge, we are the first to compare and contrast how PWYW and NYOP perform relative to each other as a function of the environment.

Research on Name Your Own Price has been strongly influenced by the business model of Priceline. The majority of the related research focuses on NYOP sellers' design decisions such as repeated bidding and bidding fees (Fay, 2004; Spann, Zeithammer, & Häubl, 2010), joint bidding for multiple items (Amaldoss & Jain, 2008), bid/price elicitation (Chernev, 2003; Spann, Häubl, Skiera, & Bernhardt, 2012) and haggling (Terwiesch, Savin, & Hann, 2005). In addition, there are some papers studying the effects of competition on the profitability of a NYOP channel (Fay, 2009; Shapiro, 2011) and reasons for the existence of the channel itself (Wang, Gal-Or, & Chatterjee, 2009).

Another stream of research on NYOP is concerned with buyers' bidding behavior and related papers analyze the role of bidders' emotions (Ding, Eliashberg, Huber, & Saini, 2005), expectations about changes in sellers' threshold level (Fay & Laran, 2009; Fay & Lee, 2015), information diffusion about seller's threshold level (Hinz & Spann, 2008) and adaptability of the threshold level (Hinz, Hann, & Spann, 2011) on buyers' bidding behavior. Further, empirical research using historic NYOP bidding data analyzes bidder characteristics such as frictional costs (Hann & Terwiesch, 2003; Terwiesch et al., 2005), rationality (Spann & Tellis, 2006), risk aversion (Abbas & Hann, 2010) or willingness to pay (Spann, Skiera, & Schäfers, 2004).

The remainder of this chapter is organized as follows: In Section 2.2, we describe the experimental design and procedures of our first set of experiments. Section 2.3 provides theoretical predictions for seller and buyer behavior in PWYW and NYOP. The results are discussed in Section 2.4. Section 2.5 reports the experimental design and results for the second set of experiments, in which one particular channel of additional benefits (i.e., word of mouth) is derived endogenously and explored in more detail. Section 2.6 concludes. All formal proofs and the experimental instructions can be found in Appendix B.

2.2 EXPERIMENTAL DESIGN & PROCEDURES

2.2.1 *General Setup*

To compare the functioning and the performance of PWYW and NYOP to traditional posted prices we consider three pairs of treatments.

1. In the first two treatments two sellers compete for customers. The first seller has to quote a posted price. The second seller can choose between quoting a posted price and delegating the pricing decision to customers. In treatment PCFLEX (*P*WYW, *C*ompetition, *Flex*ible role) the flexible seller can opt for PWYW, in treatment NCFLEX (*N*YOP, *C*ompetition, *Flex*ible role) he can opt for NYOP. With these treatments we analyze under what conditions sellers choose to use customer-driven pricing mechanisms if they have to compete against traditional posted-price sellers.

2. In two control treatments one seller again has to quote a posted price while the other seller is now constrained to either using PWYW (in treatment PCFIX) or to using NYOP (in treatment NCFIX). Here the question is not whether the customer-driven pricing mechanism is preferred to quoting a posted price, but rather under what conditions these mechanisms are viable in a competitive environment.

3. Finally, we conducted two monopoly treatments in which there is just one seller. In treatment PM (*P*WYW, *M*onopoly) the monopolistic seller has to use PWYW, in NM (*N*YOP, *M*onopoly) he has to use NYOP. These treatments show how much customers are willing to pay voluntarily in the absence of alternative suppliers. They are also interesting in their own right because there are some markets with very little competition in which customer-driven pricing mechanisms are frequently used (e.g., museums, churches, etc.).

At the beginning of each session instructions are read aloud. Then subjects have to answer a set of control questions before the experiment starts. At the end of each session we elicit information about risk preferences and social preferences of the participants and their demographic characteristics.[10] In each treatment, subjects are randomly assigned to a

10 For risk preferences we use a menu of ten paired lottery choices adapted from Holt and Laury (2002). For social preferences we rely on the six primary social value orientation

role (i.e., buyer or seller) that remains fixed throughout the experiment. Each session consists of 24 subjects which gives us three markets in the competition treatments (two sellers facing six buyers in each market) and six markets in the monopoly treatments (one seller facing three buyers). All treatments are repeated for 20 periods, and subjects are randomly rematched every period. We conduct eight sessions of the competition treatments and another eight sessions of the monopoly treatments. In order to perfectly control the valuations of the buyers and the cost of the sellers we use an induced-value design (Smith, 1976).

A novel feature of our experimental design is a per-unit benefit that may accrue to sellers using a customer-driven pricing mechanism. In this first set of experiments the benefit is exogenously given. This reduces the complexity of the experimental design and allows for both direct benefits (media buzz and word-of-mouth recommendations that increase the customer base) and indirect benefits (promoting complementary products of the same seller). In a second set of experiments, reported in Section 2.5, we generate direct promotional benefits endogenously in the lab.

The benefit $b \in \{0, \overline{b}\}$ is proportional to the number of units sold. In order to identify the effect of b we assigned $\overline{b}$ randomly to 50 percent of all markets. Sellers know whether they enjoy a positive benefit from using a customer-driven pricing mechanism while customers know only that these benefits exist with probability 0.5. This design choice reflects the fact that most buyers do not know how large the positive external effects on sellers using customer-driven pricing mechanisms are, while they may have better idea of the marginal cost of production.[11]

A total of 384 subjects participated in the experiment, 192 in the monopoly treatments and 192 in the competition treatments. Sessions lasted about two hours and subjects earned on average 18 Euros (about 24 US Dollars at the time of the experiment), including a show-up fee

(SVO) slider items of Murphy, Ackermann, and Handgraaf (2011). The SVO measure consists of a series of allocation decisions that can be used to classify the social preferences of the decision maker. At the end of the experiment, one randomly selected decision of the Holt & Laury-task and one randomly selected allocation decision from the SVO measure are paid out.

11 This argument rests on the assumption that customers may infer costs from the degree of competition within a given industry, which should provide an indication for the size of the markup charged by sellers. Alternatively, prices during sales periods may serve to narrow down the range of potential actual costs, effectively revealing an upper bound.

of 4 Euros.[12] All sessions were conducted at the experimental laboratory of the University of Munich (MELESSA). The subject pool consisted mainly of students from a wide range of majors. Treatments were implemented using zTree (Fischbacher, 2007) and subjects were recruited using ORSEE (Greiner, 2004).

2.2.2 *Competition Treatments*

COMPETITION WITH FLEXIBLE ROLES In treatments PCFLEX and NCFLEX one of the two sellers can choose whether to use posted prices or to use PWYW (NYOP, respectively) while the other (traditional) seller has to use a posted price. At the beginning of each period all subjects observe the per-unit production cost of the good which is the same for both sellers and drawn from $c \in \{10, 30, 50\}$. The flexible seller privately learns the per-unit benefit $b \in \{0, 40\}$ from using a customer-driven pricing mechanism and each buyer privately learns his valuation of the good which is drawn independently from $v \in \{10, 25, 40, 60, 120, 200\}$.[13] Then each seller decides whether to enter the market, and the flexible seller decides which pricing method to use. Thereafter all buyers and sellers are informed about the market structure. Now the posted-price sellers set their prices, and a NYOP seller sets the (secret) threshold above which all price offers are accepted. Finally buyers decide whether and if so from which seller to buy. If they go for a posted-price seller they have to pay the posted price. If they go for a PWYW seller they get the good with certainty and can choose how much to pay for it voluntarily (including a price of zero). If they go for a NYOP seller they submit a bid. If the bid is greater than or equal to the secret threshold set by the seller, they pay their bid and receive the good. If the bid is smaller than the threshold, they do not receive it and do not have to pay. Finally, payoffs are made.

12 In all competition treatments, subjects received an additional payment for completing a survey at the end of the experiment. This payment was announced upon completion of the main experiment, so that the decisions in the experiment cannot be distorted by income effects.

13 The timing of uncertainty resolution with respect to the benefit implies that sellers can perfectly anticipate whether there is a benefit or not. One interpretation of this design feature is that sellers know about their users' activities in social networks, i.e. whether they are highly engaged or rather passive. Sellers with a highly engaged user community should therefore expect that their consumers will be more likely to spread positive word of mouth. As a consequence, these sellers will be able to acquire more new customers.

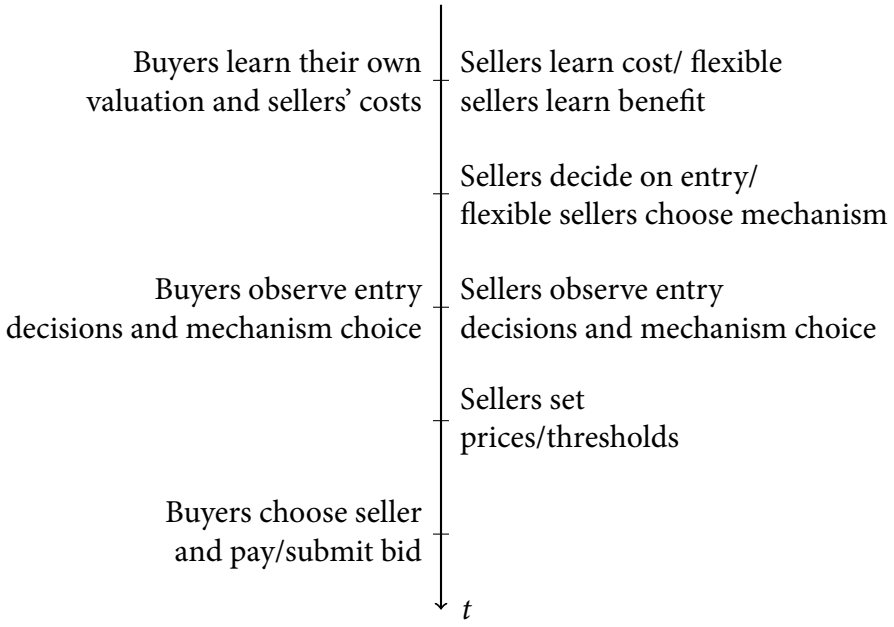


FIGURE 2.1.—Sequence of Events Flexible Competition

Figure 2.1 summarizes the time and information structure of the competition treatments with flexible roles.

A seller who stays out of the market gets a payoff of zero. If seller j with cost c quotes a posted price p^P his payoff is given by

$$\pi_j^P = \sum_{i=1}^{6} \mathbf{1}_i^j [p^P - c], \tag{2.1}$$

where $\mathbf{1}_i^j$ is an indicator function equal to one if buyer i decides to buy from seller j. If seller j uses PWYW/NYOP his payoff is

$$\pi_j^{\text{PWYW/NYOP}} = \sum_{i=1}^{6} \mathbf{1}_i^j [p_i - c + b], \tag{2.2}$$

where p_i refers to the price paid in PWYW and the submitted bid in NYOP, respectively, and b is the per-unit benefit. Here, $\mathbf{1}_i^j$ is equal to one if buyer i decides to buy from seller j (and his submitted bid is greater than or equal to the threshold in the case of NYOP) and zero otherwise. The payoff of buyer i is given by $v_i - p_i$ if a transaction takes place, and zero otherwise, where v_i is his valuation and p_i the price he paid.

COMPETITION WITH FIXED ROLES The competition treatments with fixed roles (treatments PCFIX and NCFIX) are identical to the

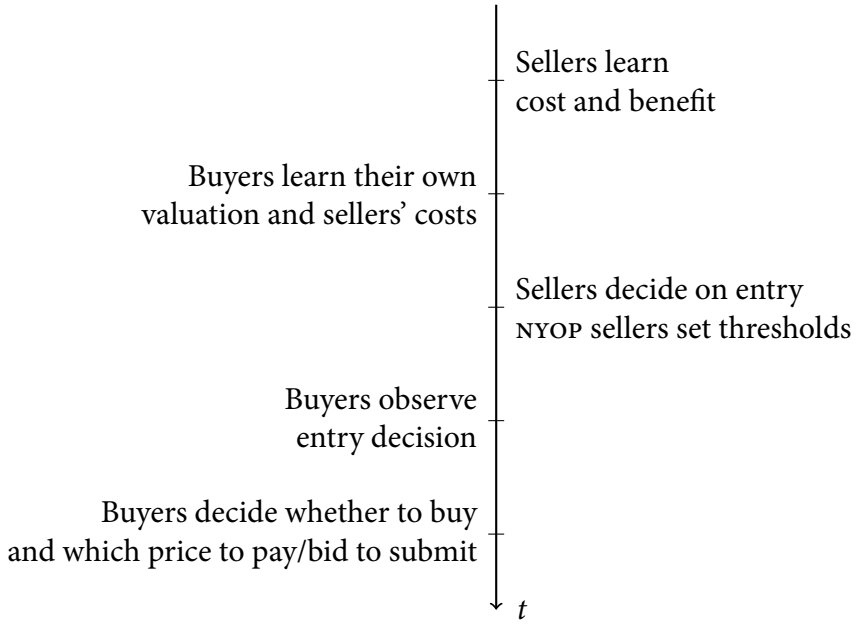


FIGURE 2.2.—Sequence of Events Monopoly

treatments with flexible roles except for the fact that one of the sellers has to use either PWYW or NYOP if he enters the market.

2.2.3 *Monopoly Treatments*

In two monopoly treatments (PM and NM) there is only one seller who is forced to use PWYW (NYOP, respectively) if he enters the market. Here there are only three buyers in each market. Costs and benefits are parameterized as in the competition treatments, while buyers' valuations are drawn from a restricted set, $v = \{40, 60, 120, 200\}$.[14] All sellers have the same cost-benefit combination in a given period as in the competition treatments and learn that combination before market entry. The time structure of the monopoly treatments is depicted in Figure 2.2.

Seller j's payoff is given by

14 The competition treatments comprise a larger market and contain the additional valuations of 10 and 25. Therefore, profits cannot be directly compared between monopoly and competition treatments. However, in the monopoly treatments we are mainly interested in the behavior of buyers. For each given valuation of a buyer we can compare the behavior in the monopoly and the competition treatments.

$$\pi_j = \begin{cases} \sum_{i=1}^{3} \mathbf{1}_i[p_i - c + b], & \text{if he entered the market} \\ 0, & \text{if he did not enter the market.} \end{cases} \tag{2.3}$$

Again, $\mathbf{1}_i$ is an indicator function equal to one if buyer i decides to buy (and his submitted bid is greater than or equal to the threshold in the case of NYOP) and zero otherwise. The payoff of buyer i is again $v_i - p_i$.

2.3 THEORETICAL PREDICTIONS

As a benchmark for the experimental results we analyze the experimental treatments under the assumption that all buyers and sellers are fully rational and purely self-interested. To avoid uninteresting case distinctions we treat prices as continuous variables. Furthermore, we start out with the assumption that all subjects are risk neutral. The effects of risk aversion are discussed at the end of this section.

We solve the experimental games by backward induction. In the case of PWYW the analysis is straightforward:

Proposition 2.1. *A purely self-interested buyer facing a PWYW seller pays $p = 0$ and gets a strictly positive payoff $\pi_i = v_i - 0 > 0$. In all three PWYW treatments all buyers buy from the PWYW seller if this seller entered the market.*

- *In PCFIX and PM the PWYW seller enters the market if and only if his benefit exceeds his cost, i.e., $b = 40$ and $c \in \{10, 30\}$. In this case he captures the entire market and makes a strictly positive profit while the PP seller makes a profit of zero. If the PWYW seller does not enter, the PP seller charges the monopoly price.*
- *In PCFLEX the flexible seller always enters the market and chooses PWYW if and only if his benefit exceeds his cost, i.e., $b = 40$ and $c \in \{10, 30\}$. In this case he captures the entire market and makes a strictly positive profit. If the flexible seller chooses to offer a posted price, both sellers engage in Bertrand competition and make zero profits.*

Proposition 2.1 gives rise to the following prediction:

Prediction 2.1. *In all three PWYW treatments sellers use the PWYW pricing mechanism if and only if $b = 40$ and $c \in \{10, 30\}$. If PWYW is offered, then it captures the entire market, but all buyers pay a price of 0. The performance of PWYW is independent of competition.*

These results suggest that the additional benefits generated by PWYW (e.g., via word-of-mouth effects) are the only reason to use this mechanism. However, these results are based on the assumption that all buyers are purely self-interested. If some buyers have social preferences and are motivated by fairness or reciprocity, they may be willing to pay positive prices voluntarily. Furthermore, if a buyer feels obliged to make a positive payment under PWYW but does not want to engage in moral deliberations about how much to pay, he may prefer to buy from a posted-price seller, if such a seller is available.

Let us now turn to the NYOP treatments. In these treatments we ignore the entry decision because sellers could always guarantee non-negative profits if they entered. In fact almost all sellers choose to enter in the NYOP treatments.[15] If a seller uses the NYOP mechanism he has to set a secret threshold t. A buyer gets the good if and only if the price he bids exceeds this threshold, i.e., $p \geq t$. The following lemma shows how a NYOP seller optimally sets this threshold:

Lemma 2.1. *If a seller uses NYOP it is a (weakly) dominant strategy to set* $t^* = \max\{c - b, 0\}$ *in all treatments.*

Proof. See Appendix A. □

The intuition for this result is straightforward. The seller's "effective" marginal cost is $c - b$, and he cannot set a threshold smaller than zero. Clearly, it cannot be optimal to set the threshold below his effective marginal cost. Setting the threshold strictly above effective marginal cost cannot be optimal either, because the threshold is not observed by customers and cannot affect their behavior. Thus, the only effect of a threshold greater than the effective marginal costs is to reject some profitable price offers.

In the following we restrict attention to Perfect Bayesian equilibria in which a seller using the NYOP mechanism always chooses his (weakly) dominant strategy.[16]

15 The fraction of sellers entering the market is above 98 percent in each of the three NYOP treatments.

16 In the monopoly treatment this assumption is without loss of generality because $t = t^*$ is strictly optimal. In the duopoly treatments there are Perfect Bayesian equilibria in which the NYOP seller chooses a threshold $\bar{t} > c$, the posted-price seller offers p^P such that $c \leq p^P < \bar{t}$ and all buyers with $v_i > p^P$ buy from the posted-price seller. These equilibria are sustained by the beliefs of the buyers that the NYOP seller sets the high threshold $\bar{t} > c$, so they never buy from this seller. Given that they do not buy, the thresthold $\bar{t}$ is (weakly) optimal. However, these equilibria are not trembling hand perfect. If there is an arbitrarily small probability that a buyer makes an offer to the NYOP seller, it is strictly optimal for the NYOP seller to set $t = t^*$.

Lemma 2.2. *In the competition treatments a posted-price seller chooses $p^P \geq c$ and makes zero profits in equilibrium.*

Proof. See Appendix A. □

The intuition is again straightforward. Either both sellers use posted prices, in which case Bertrand competition drives prices down to marginal costs and profits to zero, or the flexible seller uses NYOP. Given his optimal threshold consumers can always get the good by bidding c. Thus, the posted-price seller may either set $p = c$ (in which case he may capture some market share) or $p > c$ in which case all consumers shop with the NYOP seller, but the posted-price seller's profits are zero in any case.

What is the optimal behavior of a risk neutral and self-interested buyer? The buyer rationally anticipates that the NYOP seller chooses threshold $t^* = \max\{c - b, 0\}$. In NCFIX and NM the buyer knows that the actual benefit is $b = 40$ or $b = 0$ with equal probability.

Lemma 2.3. *In treatments NCFIX and NM a risk neutral buyer facing a NYOP seller offers*

$$p = \begin{cases} c & \text{if } v \geq \min\{2c, c + 40\} \\ \max\{c - 40, 0\} & \text{if } v < \min\{2c, c + 40\} \end{cases} \tag{2.4}$$

Proof. See Appendix A. □

The proof follows directly from Lemma 2.2 and the buyer's payoff function. Note that the buyer will bid aggressively only if his valuation is sufficiently small. Table 2.1 displays the optimal prices offered by a risk-neutral buyer to a NYOP seller as a function of the values that v and c could take in the experiment.

Boldfaced values indicate cases in which a risk neutral buyer optimally submits an aggressive, "risky" bid to the NYOP seller that is strictly smaller than the seller's cost. In these cases it is strictly optimal for a risk neutral buyer to buy from the NYOP seller. If, on the other hand, a safe bid of $p = c$ is optimal, then the buyer may also buy from the posted-price seller, provided that this seller charges $p^P = c$.

In NCFLEX the analysis is slightly more complicated. In this treatment the seller chooses whether to use NYOP after privately observing the realization of b. Thus, his choice may signal information about the realization of b to the buyer. There exists a pooling equilibrium in which the flexible seller always uses NYOP and the buyer offers the prices given by Lemma 2.3. However, there is also a separating equilibrium in which

	Valuation					
Cost	$v = 10$	$v = 25$	$v = 40$	$v = 60$	$v = 120$	$v = 200$
$c = 10$	**0**	10	10	10	10	10
$c = 30$	**0**	**0**	**0**	30	30	30
$c = 50$	**10**	**10**	**10**	**10**	50	50

TABLE 2.1.—Optimal Bidding Behavior

Notes: Table displays optimal bids for all cost-valuation combinations, under the assumption of risk neutrality. Boldfaced values indicate cases in which a risk neutral buyer optimally submits a bid to the NYOP seller that is strictly smaller than the seller's cost.

the seller uses NYOP if and only if $b = 40$. Thus, if the buyers observe that NYOP is offered, they conclude in equilibrium that $b = 40$ and offer $p = \max\{c - 40, 0\}$ for all realizations of v. This separating equilibrium is less profitable for the seller than the pooling equilibrium because prices and profits are lower, so sellers prefer the pooling equilibrium. Nevertheless, we cannot rule out the possibility that buyers interpret NYOP as a signal that $b = 40$ and therefore bid more aggressively than they do in NCFIX. But even in this case the NYOP seller makes strictly positive profits if $40 - c > 0$.

This analysis is summarized in the following prediction for the NYOP treatments:

Prediction 2.2.

1. *In all three NYOP treatments a NYOP seller sets the optimal threshold* $t^* = \max\{c - b, 0\}$. *In the competition treatments a posted-price seller sets* $p^P \geq c$ *in equilibrium.*

2. *Buyers with a low valuation* ($v < \min\{2c, c + 40\}$) *buy from the NYOP seller and offer* $p = \max\{c - 40, 0\}$ *which is successful in at least* 50 *percent of all cases. If buyers with a higher valuation buy from the NYOP seller, they offer either* $p = \max\{c - 40, 0\}$ *or* $p = c$ *which is successful with probability one in equilibrium. They may also buy from the posted-price seller, but only if the posted-price seller offers* $p^P = c$.

3. *NYOP sellers make positive profits on average, while posted-price sellers always make zero profits in equilibrium. Furthermore, NYOP*

sellers always have a higher expected market share than posted-price sellers. If there is competition, the seller prefers NYOP *over posted prices.*

These predictions rest on several assumptions. First, we assumed that all subjects are rational and that rationality is common knowledge. Thus, buyers correctly anticipate the optimal threshold set by the NYOP seller. However, it is possible that some NYOP sellers fail to understand what the optimal threshold is and choose a threshold that is too high. Even if they choose the threshold optimally, buyers may believe that sellers behave irrationally with some probability. This may induce buyers to offer prices that are higher than the optimal threshold. It may also induce buyers to buy from the posted-price seller where they get the good with certainty. Finally, buyers may not be fully rational and have difficulties to compute the optimal threshold themselves. Again, this may induce them to offer a price that is higher than the seller's cost or to choose the posted-price seller.

Second, we assumed that buyers are risk neutral. If a buyer is risk averse, he may prefer a "safe" offer $p = c$ even if the expected profit of a "risky" offer $p = \max\{c - 40, 0\}$ is higher. Furthermore, if there is no common knowledge of rationality, risk aversion will exacerbate the effects of strategic uncertainty discussed above, i.e., the buyer may be inclined to offer even higher prices to the NYOP seller or lean more toward the posted-price seller where he can get the good with certainty.

Finally, we assumed that buyers are purely self-interested. If a buyer cares about the utility of the seller, he may offer higher prices to the NYOP seller in order to achieve a more equal income distribution.

2.4 RESULTS

We first analyze under what circumstances sellers choose to employ a customer-driven pricing mechanism and how successful these mechanisms are in penetrating the market and in making profits. Then, we analyze how buyers react to these mechanisms. Do they buy from a PWYW or NYOP seller or do they shy away from them? What prices do they pay? Do they behave differently if there is another posted-price seller as compared to a situation where the PWYW or NYOP seller is a monopolist? Finally, we analyze the competitive effects of sellers employing PWYW and NYOP relative to a posted price competitor.

2.4.1 *Performance of Customer-Driven Pricing Mechanisms*

Under what conditions do sellers choose to delegate pricing power to their customers?

Result 2.1 (Seller's Choice of Customer-Driven Pricing Mechanism)**.**

(a) In PCFlex almost all sellers choose to use PWYW if and only if PWYW offers an additional benefit and if production costs are not too high. If PWYW is chosen, then the PWYW seller captures almost all of the market and makes high profits, while the PP seller makes profits close to zero.

(b) In NCFlex almost all sellers choose to use NYOP if NYOP offers an additional benefit. If there is no benefit still about half of the sellers choose NYOP if costs are not too high. If NYOP is chosen then the NYOP seller captures about 60 percent of the market. If NYOP is chosen, it is significantly more profitable than quoting a posted price. NYOP sellers choose a threshold close to the optimal threshold in 75 percent of all cases.

(c) A flexible seller who chooses PWYW makes profits that are more than twice as high as the profits made by a flexible seller who chooses NYOP.

Support for Result 2.1 is provided by the descriptive statistics reported below. Table 2.2 reports the percentage of cases in which sellers opt for one of the customer-driven pricing mechanisms by cost and benefit levels. In both treatments all flexible sellers entered the market.

In PCFlex almost all sellers shy away from PWYW if $b = 0$. If there is a benefit and if $b > c$ all sellers choose PWYW, and 13 percent do so if $c = 50$ and $b = 40$. This result confirms Prediction 2.1. It seems that sellers are convinced that buyers are not going to make voluntary payments, so sellers avoid PWYW if this may result in losses. Note, however, that in the case where $c = 50$ and $b = 40$ there are twice as many sellers offering PWYW than if $c = 10$ and $b = 0$, even though the "effective" marginal cost is the same. This suggests that sellers expect buyers to pay more if their costs are high than if their costs are low.

In NCFlex almost all flexible sellers choose NYOP if there is a positive benefit. If there is no benefit, still almost 50 percent of the sellers opt for NYOP if costs are 10 or 30, and 17 percent do so if costs are 50. This result is consistent with Prediction 2.2. Furthermore, it suggests that the

	PCFlex		NCFlex	
Cost	$b = 0$	$b = 40$	$b = 0$	$b = 40$
$c = 10$	6%	100%	50%	100%
$c = 30$	0%	100%	43%	94%
$c = 50$	0%	13%	17%	100%
Total	2%	65%	40%	98%

Table 2.2.—Pricing Mechanism Choice

Notes: Entries in cells denote the percentage of cases in which sellers opt for one of the customer-driven pricing mechanisms, conditional on entering the market.

choice of NYOP could be interpreted as a signal that the seller is more likely to enjoy a high benefit.

Using PWYW is highly profitable. The average profit of a PWYW seller is 127.1 points, as compared to a profit of almost zero of the competing PP seller. NYOP is somewhat less profitable with an average profit of 58.1 which is still much higher than the average profit of 15.7 of the corresponding PP seller (see Table 2.3). However, because sellers could choose whether or not to employ the customer-driven pricing mechanism, these numbers have to be interpreted with caution. In PCFlex sellers opted for PWYW only if the benefit was high, while in NCFlex many sellers also chose NYOP when there was no benefit. But even if we restrict attention to the cases with $b = 40$ and $c \in \{10, 30\}$ where almost all flexible sellers opted for a customer-driven pricing mechanism the average profit under PWYW is 145.1 while NYOP sellers make only 95.6 on average.

The reason why PWYW is more profitable than NYOP if benefits are higher than costs is the fact that PWYW is much more successful in market penetration. If the flexible seller chooses PWYW he gets a market share of 94.2 percent.[17] If the flexible seller chooses NYOP his market share is only 57.4 percent.[18]

The overall picture of PWYW does not change much when we look at treatments PCFix and PM in which one seller had to use PWYW. Note that this seller could still decide not to enter the market. In fact, in PCFix the PWYW seller entered in only 51.7 percent of all cases (53.8 percent in PM). He stayed out of the market when there was no benefit

17 The PP seller gets only 3.3 percent of the market, 2.5 percent of the buyers do not buy.
18 The PP seller gets 35.5 percent of the market and 7.1 percent of the buyers do not buy.

	Profits				Market Shares	
Treatment	PWYW	PP	NYOP	PP	PWYW	NYOP
Flexible Competition	127.1	0.9	58.1	15.7	94.2%	57.4%
Fixed Competition	76.8	4.7	69.7	22.4	90.0%	69.8%
Monopoly	52.7		83.2		98.6%	97.8%

Table 2.3.—Profits and Market Shares

Notes: Profits and market shares are conditional on market entry of both sellers and choice of the customer-driven pricing mechanism. Both profits and market shares are displayed as per period averages.

($b = 0$) and/or his cost was high ($c = 50$), so in the same situations in which he opted for posted prices in PCFlex. If he entered, he again captured almost the entire market (market share 90.0 percent). If $b = 40$ and $c \in \{10, 30\}$ the profits of the PWYW seller in PCFix are virtually identical to the profits in PCFlex. However, in PCFix some PWYW sellers enter the market when $b = 0$. In these cases they make losses on average, confirming the pessimistic beliefs of sellers in PCFlex. This explains why the average profit conditional on market entry is lower, see Table 2.3. It is also interesting to note that profits are negative if $b = 0$ and $c = 10$, but positive if $b = 40$ and $c = 50$, even though these two situations are strategically equivalent. Thus, it must be the case that buyers voluntarily paid higher prices if the seller had higher costs. We will get back to this question below.

NYOP sellers' profits in NCFix are not significantly different from profits in NCFlex. Moreover, we find that buyers do not submit lower bids in NCFlex as compared to NCFix. This indicates that the choice of NYOP does not provide an informative signal about the level of the benefit.

Did NYOP sellers set the thresholds optimally? By Proposition 2.1 it is a weakly dominant strategy to set $t = t^* = \max\{c - b, 0\}$. Across all treatments, 30.0 percent of the sellers choose exactly t^*, and 72.4 percent choose a threshold within 10 points of t^*. There is no significant difference between treatments. On average the actually chosen threshold is 8.7 points higher than the optimal threshold, but this difference is

decreasing over time suggesting that sellers learn to set the threshold optimally as they gain experience.[19]

2.4.2 *Buyers' Reactions to Customer-Driven Pricing Mechanisms*

How did buyers react if they were offered customer-driven pricing mechanisms?

PAY WHAT YOU WANT

Result 2.2 (Voluntary Payments under PWYW)**.**

(a) Almost all buyers buy from a PWYW seller.

(b) The majority of buyers (56.2 percent) pay positive prices and 26.2 percent pay prices greater than or equal to the seller's cost. On average each buyer pays 9.8 points, which is a significant contribution to sellers' profits.

(c) Buyers tend to pay more the higher their valuation and the higher the seller's cost. Payments are higher the higher the social value orientation (SVO) of a buyer. Buyers also pay more if there is no competing posted-price seller. Voluntary payments tend to decrease over time.

The high market share of more than 90 percent of PWYW sellers shows that buyers do not hesitate to buy from a PWYW seller. The PWYW market share is higher compared to previous research which finds that there is some fraction of buyers who prefer not buying from a PWYW seller. This has been explained by a reluctancy to engage in the moral deliberations of how much to pay (Schmidt et al., 2015), self signaling (A. Gneezy, Gneezy, Riener, & Nelson, 2012) or privacy concerns (Regner & Riener, forthcoming).

In the current study, however, this fraction is lower presumably because of the additional benefits that PWYW sellers enjoy. Buyers know that a seller who chooses to enter the market and offers PWYW is likely to enjoy this benefit, so the buyer does not have to feel bad about accepting a PWYW offer even if he pays less than the seller's cost. Thus, the

19 A random-effects regression of the chosen threshold on the optimal threshold shows that t^* explains the actually chosen threshold very well. The coefficient of t^* is 0.9. The treatment dummies and the interaction effects are not significant, but there is a significant negative time trend.

additional benefit helps to make PWYW a very effective instrument for market penetration.

Table 2.4 reports the average prices paid as well as the fractions of buyers paying positive prices and prices greater than or equal to the seller's cost in all three PWYW treatments. These voluntary payments are a significant contribution to the sellers' profits under PWYW, contradicting Proposition 2.1. Table 2.4 shows that social preferences do play an important role.

	PCFlex	PCFix	PM	Total
Average price paid	4.9	7.6	12.1	9.8
Fraction of buyers paying $p > 0$	38.1%	50.4%	64.1%	56.2%
Fraction of buyers paying $p \geq c$	15.9%	19.4%	32.2%	26.2%

TABLE 2.4.—Prices Paid under PWYW

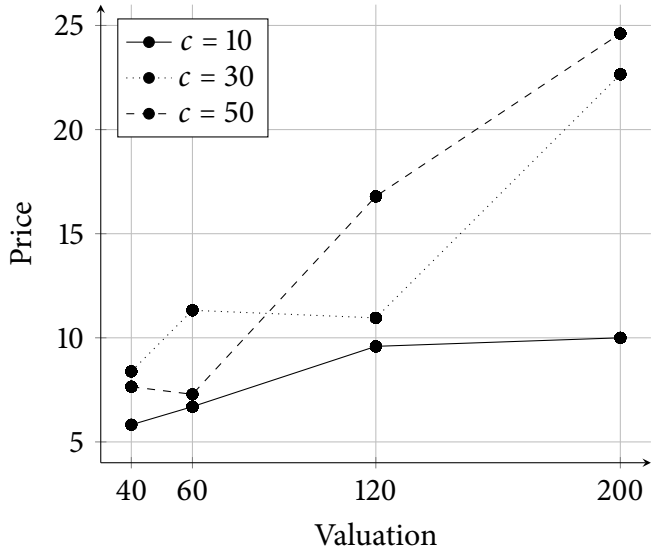
Prices are lowest in PCFlex, higher in PCFix and highest in PM. A possible explanation is that in PCFlex sellers had the outside option to choose posted prices which they almost always did if $b = 0$ and/or $c = 50$. Thus, in PCFlex PWYW was profitable even if no positive prices were paid. In contrast, in PCFix and PM the outside option was to stay out of the market, so sellers choose to enter with PWYW more often, even in situations in which they could make losses. This may have induced buyers with social preferences to be more generous. Furthermore, in the monopoly treatment buyers had more reason to be grateful if the PWYW seller entered the market. In the monopoly treatment, if the seller did not enter, buyers got a payoff of zero. In the competition treatment they could still buy from the PP seller and get positive payoffs. This is a possible explanation for the higher payments observed in the monopoly treatment. Finally, valuations are on average lower in PCFix and PCFlex, due to the addition of a low-valuation market segment in these latter treatments. Therefore we should expect average PWYW prices to be lower as well.

The regressions displayed in Table 2.5 show the driving forces of the behavior of buyers. Prices paid under PWYW are significantly increasing in the buyer's valuation and in the seller's cost. Thus, PWYW achieves endogenous price discrimination, consistent with models of social preferences. Panel 2.3a visualizes the increasing relationship between valuations and prices for all cost levels.

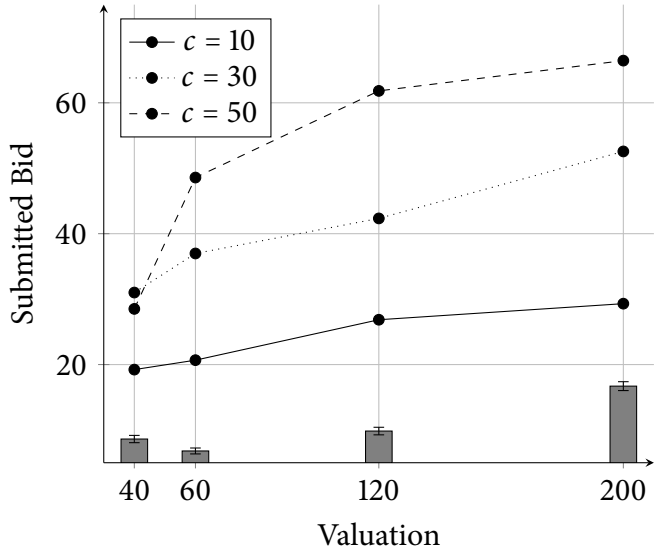
	(1) Price Paid in PCFlex	(2) Price Paid in PCFix	(3) Price Paid in PM
Cost	0.058	0.212***	0.208***
	(0.094)	(0.061)	(0.037)
Valuation	0.092***	0.117***	0.102***
	(0.015)	(0.014)	(0.009)
Posted Price of S^P	-0.001	-0.015	
	(0.026)	(0.027)	
Period	-0.586***	-0.650***	-0.359***
	(0.206)	(0.170)	(0.112)
SVO	0.421	0.638***	0.606***
	(0.317)	(0.227)	(0.133)
Constant	-19.701**	-19.139***	-18.431***
	(8.512)	(6.493)	(3.848)
Number of observations	226	335	763

TABLE 2.5.—Determinants of Behavior (PWYW)

Notes: Random-effects tobit regressions on buyers' prices paid to the PWYW seller (left-censored at minimum price of 0). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses.



(A) PWYW



(B) NYOP

FIGURE 2.3.—Endogenous Price Discrimination

Notes: The upper panel displays average per unit prices paid to PWYW sellers by cost levels. The lower panel displays average submitted bids to NYOP sellers by cost levels. In addition, bars show the extent of overbidding relative to the theoretical optimum by valuation, averaged over all cost levels (assuming common knowledge of rationality and risk neutrality).

The importance of social preferences is confirmed by the fact that social value orientation (SVO) also has a highly significant impact.[20] The price of the competing posted-price seller in the competition treatments has no significant effect. Note that the variable "Period" has a significantly negative coefficient. This suggests that PWYW is more successful if it is newly introduced and if buyers have not yet gotten used to it.

NAME YOUR OWN PRICE

Result 2.3 (Bidding Behavior under NYOP)**.**

(a) If NYOP is offered under competitive conditions about 60 percent of all buyers choose the NYOP seller.

(b) Buyers are more likely to choose NYOP the lower their valuation, the lower the seller's cost and the higher the posted price of the competing seller.

(c) Most buyers submit bids that are significantly higher than the optimal threshold of the seller. On average they are also significantly higher than the actual thresholds chosen by sellers. Bids tend to increase with the valuation of the buyer.

NYOP is significantly less successful in market penetration than PWYW (see Table 2.3). Table 2.6 reports average bid amounts and the fraction of successful bids in all three NYOP treatments. Bids are higher in the monopoly treatment than in the competition treatments.

	NCFLEX	NCFIX	NM	Total
Average bid	21.8	23.1	39.5	33.5
Fraction of successful bids	59.8%	61.4%	83.5%	75.5%

TABLE 2.6.—Bids Submitted under NYOP

Regression (1) in Table 2.7 offers some insights into which buyers go for NYOP under which circumstances. The regression shows that buyers are less likely to choose a NYOP seller if their valuation is high and if the posted price of the competing seller is low. This is very intuitive.

20 In PCFLEX the seller's cost and the buyer's SVO are not significant. This could be related to the fact that a seller's mechanism choice signals to buyers that the benefit exceeds the marginal cost, and hence PWYW is profitable even if buyers pay a price of zero.

Given that not all NYOP sellers set the threshold optimally there is some strategic uncertainty whether a buyer gets the good under NYOP while he can be sure to get it if he pays the posted price of the competing seller. Thus, the higher his valuation (i.e., the more is at risk) and the lower the competing posted price the more is a buyer inclined not to buy from a NYOP seller. Surprisingly, our measure of risk aversion does not have a significant impact on buyers' choices. The buyers' social value orientation (SVO) is marginally significant (at the 10 percent level) suggesting that more socially minded buyers tend to use NYOP less often. The significant positive time trend suggests that buyers become more prone to use NYOP as they get more experienced.

Regressions (2) to (4) in Table 2.7 show the driving forces of the bids submitted if a buyer chooses the NYOP seller. The theoretically optimal bid derived under the assumption of common knowledge of rationality and risk neutrality (see Table 2.1) is highly significant in the NCFIX and NM treatments. According to Prediction 2.2 the coefficients of "Optimal Bid" should be equal to 1, but they are much smaller.[21] The buyer's valuation also has a highly significant positive effect. Again, this is very intuitive. Given that many sellers choose thresholds that are too high, buyers cannot be sure to get the good if they submit the theoretically optimal bid. Thus they will bid higher the more is at stake. We do not find a significant effect of our measure of risk aversion and of SVO. However, Panel 2.3b lends descriptive support to the hypothesis that risk aversion is the more important determinant of bids: since overbidding relative to the theoretical optimum under risk neutrality increases (almost) monotonically in valuation, individuals seem to become more risk averse the higher the potential for losses. There is a significant negative time trend suggesting that buyers bid more aggressively as they gain more experience in NM.

2.4.3 *Customer-Driven Pricing Mechanisms as a Competitive Strategy*

How do customer-driven pricing mechanisms affect competition? Figure 2.4 compares profits of the seller using PWYW or NYOP to the profits of his posted price competitor. In Section 2.4.1 we have already seen that if PWYW is used it is highly profitable and captures almost the entire market. Thus, not surprisingly, the profits of the competing posted-price

21 The parameter estimates for Optimal Bid are significantly different from 1 for specifications (2)–(4) in Table 2.7. (2): $\chi^2(1) = 104.84$, $p < 0.001$; (3): $\chi^2(1) = 128.83$, $p < 0.001$; (4): $\chi^2(1) = 181.99$, $p < 0.001$.

	(1) Choice of NYOP seller	(2) Buyer's bid in NCFLEX	(3) Buyer's bid in NCFIX	(4) Buyer's bid in NM
Cost	-0.0447***			
	(0.00838)			
Optimal Bid		0.0701	0.320***	0.686***
		(0.0689)	(0.0557)	(0.0222)
Posted Price of S^P	0.0683***	0.342***	0.0929***	
	(0.00835)	(0.0407)	(0.0162)	
Valuation	-0.0223***	0.189***	0.186***	0.0402***
	(0.00164)	(0.0191)	(0.0151)	(0.00597)
Risk Aversion	0.103	0.0319	0.456	-0.135
	(0.0699)	(0.284)	(0.437)	(0.369)
SVO Angle	-0.0191*	0.0658	0.0802	0.00148
	(0.0107)	(0.0495)	(0.0562)	(0.0480)
Period	0.0612***	0.130	-0.111	-0.206***
	(0.0146)	(0.0944)	(0.0844)	(0.0528)
Constant	-0.489	-3.670	2.964	23.69***
	(0.602)	(2.922)	(3.700)	(2.250)
Observations	1175	286	489	1409

TABLE 2.7.—Determinants of Behavior (NYOP)

Notes: Entries in column (1) are point estimates from a random-effects logistic regression on buyers' choice of seller in treatments NCFLEX and NCFIX. The dependent variable is 1 if the buyer opted for the NYOP seller and 0 otherwise. Entries in columns (2)-(6) are point estimates from random-effects tobit regressions on buyers' submitted bids to the NYOP seller (left-censored at minimum bid of 0). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses.

seller are very close to zero. If, on the other hand, the flexible seller also chooses posted prices, then both sellers share the market equally and make small profits.[22] Thus, the competing PP seller suffers if the flexible seller chooses PWYW.

Perhaps surprisingly this is not the case with NYOP. Profits of the PP seller are unaffected if the flexible seller opts for NYOP, while profits of the flexible seller go up significantly. This suggests that NYOP relaxes competition. In fact the markup charged by a PP seller competing against a NYOP seller in the two competition treatments is 18.8, which is significantly higher than the markup of 7.8 charged by PP sellers against a flexible seller who has chosen to use a posted price in NCFLEX. NYOP relaxes price competition because the NYOP seller does not quote a price. Furthermore, because NYOP is most attractive to low valuation customers, PP sellers can focus on high valuation customers and charge them a higher price.

Result 2.4 (The Effects on Competition). *PWYW is an aggressive competitive strategy driving a competing posted-price seller out of the market. This is not the case for NYOP. NYOP leaves room for competing sellers and it relaxes price competition. Both sellers are better off if one of them uses NYOP.*

It is important to note that additional benefits are an important precondition for most PWYW sellers to enter the market in a competitive situation. However, if these benefits exist, then a PWYW seller captures almost the entire market. This is different for a NYOP seller. NYOP can be profitable even if there is no additional benefit and if costs are high. Furthermore, a NYOP seller captures only about 60 to 70 percent of the market, leaving (mostly high valuation) customers for a posted-price competitor.

2.5 ENDOGENEOUS BENEFITS

So far we considered a setting in which the promotional benefits of customer-driven pricing mechanisms are exogenously given. This setup reduces the complexity of the experimental design and allows for multiple channels through which additional benefits may come about. In

22 The (combined) profit of 8.2 is higher than the theoretically predicted profit of zero. However, many other experiments have already shown that sellers set prices somewhat above marginal costs even if they are engaged in Bertrand competition. See, e.g., Dufwenberg and Gneezy (2000).

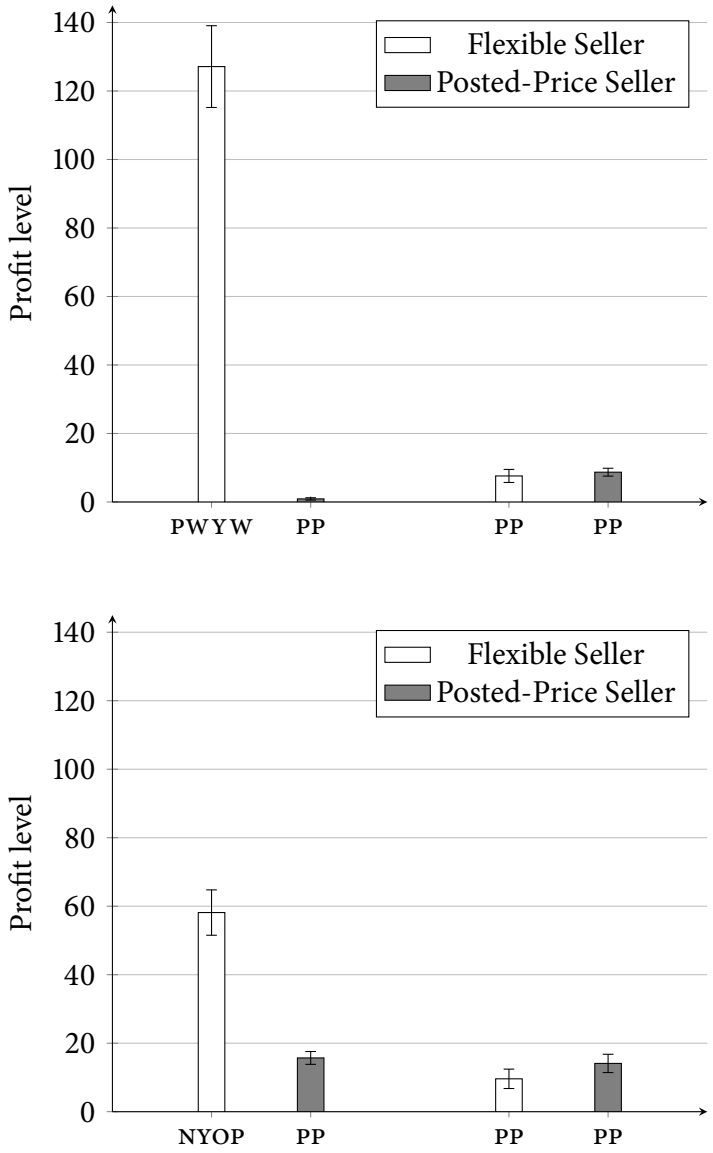


FIGURE 2.4.—Effects on Competition

Note: The upper panel shows profit levels in PCFLEX for the two different competitive environments. The two left-most bars indicate competition between a PWYW and a PP seller, while the bars on the right depict competition between two posted-price sellers. Profit levels for the NCFLEX treatment are depicted analogously in the lower panel. Throughout, the error bars show standard errors of the mean.

this section we analyze direct promotional benefits that arise if media buzz and word-of-mouth recommendations induce more consumers to buy the good. The question is under what conditions these benefits arise endogenously.

2.5.1 *Experimental Design & Procedures*

Consider the following variation of our previous design. Each market has two sellers and six buyers. However, we now split demand into two groups: There are two well-informed buyers, who are fully aware of the market structure, and four follow-up buyers, who have to rely on word of mouth to learn which sellers are in the market. We model word-of-mouth advertising in reduced-form: the purchasing decisions of the well-informed buyers directly affect the market structure for follow-up buyers. If both well-informed buyers purchase at the PWYW seller, for example, only the PWYW seller is visible to follow-up buyers. However, follow-up buyers may become fully informed about which sellers are in the market if they pay an additional search cost. Thus, there are four roles that remain fixed for the duration of the experiment: posted-price sellers, PWYW sellers (NYOP sellers, resp.), well-informed buyers and follow-up buyers.

Furthermore, in the new NYOP treatment we added the realistic feature that buyers can still buy from the posted-price seller if the bid submitted to the NYOP seller was not successful. Thus, if buyers submit a bid to a NYOP seller, they do not face the risk that they cannot consume the good if their bid fails. This should encourage more buyers to try out NYOP and to bid more aggressively.

Under the assumptions of common knowledge of rationality and risk neutrality, this should not matter for the behavior of buyers and sellers: for both sellers it is still a weakly dominant strategy to set their price or threshold equal to cost. Given prices and thresholds, well-informed buyers will therefore be indifferent between purchasing at posted prices and submitting a bid to the NYOP seller and sellers should expect to each attract three buyers in expectation.

However, if buyers are assumed to be risk averse and sellers do not compete away all seller surplus (as our first set of experiments indicate), the possibility to purchase the good from the posted-price seller if the bid submitted to the NYOP seller was not successful will induce more buyers to submit an offer to the NYOP seller. The reason is that submitting a bid is now riskless; hence, a buyer who believes that the NYOP seller

may have set a threshold lower than the competing posted price should rationally submit a bid first.

A total of 144 subjects participated in the second set of experiments, 72 in each treatment (*P*WYW *C*ompetition *E*ndogenous *B*enefit (PCEB) and *N*YOP *C*ompetition *E*ndogenous *B*enefit (NCEB)). Sessions were conducted at MELESSA. We restricted the subject pool to subjects who had not participated in any session of the first set of experiments. Sessions lasted two hours and average earnings amounted to 25 Euros (about 27 US Dollars at the time of the experiment), including a show-up fee of 4 Euros.[23]

TIMING AND INFORMATION STRUCTURE At the beginning of each period, buyers and sellers are informed about production costs ($c \in \{5, 10, 20, 30, 50\}$).[24] In addition, buyers learn their valuations ($v \in \{10, 25, 40, 60, 120, 200\}$). Then, sellers decide on entry. Upon observing the entry decisions, well-informed buyers decide whether and where to purchase.

If at least one well-informed buyer has opted for the PWYW/NYOP seller, this seller will be available for follow-up buyers at no additional cost. The same holds true for posted-price sellers. If only one seller is available to follow-up buyers, they can invest search costs ($c_s = 10$) to find out whether the other seller has entered the market; if so, they can also purchase from this seller.[25] If the market for well-informed buyers was split equally, both sellers are available at no additional cost.

After observing the market structure follow-up buyers make their purchasing decisions. If a buyer submitted an unsuccessful bid to a NYOP seller, he can turn to the posted price seller if this seller is available to him. Thus, availability of sellers for follow-up buyers in the NYOP treatment depends on the interaction between well-informed buyers and sellers. If one of the well-informed buyers submits an unsuccessful bid and subsequently purchases at the posted price, both sellers are available to follow-up buyers. If, however, both well-informed buyers interact with

23 As in the first set of experiments, treatments were implemented using zTree (Fischbacher, 2007) and subjects were recruited using ORSEE (Greiner, 2004).

24 We expected that additional benefits of using PWYW potentially exist only for low cost levels. Therefore, we kept the original cost levels from the first experiment and added two low cost levels (5 and 20) to the set.

25 To keep the PCEB and NCEB treatments symmetric, buyers have to take the decision whether or not to invest search costs before deciding where to purchase. In NCEB it is also conceivable that the buyer invests the search cost if his purchase with the NYOP seller was unsuccessful, but because there is no "unsuccessful purchase" with PWYW this cannot be implemented in the PCEB treatment.

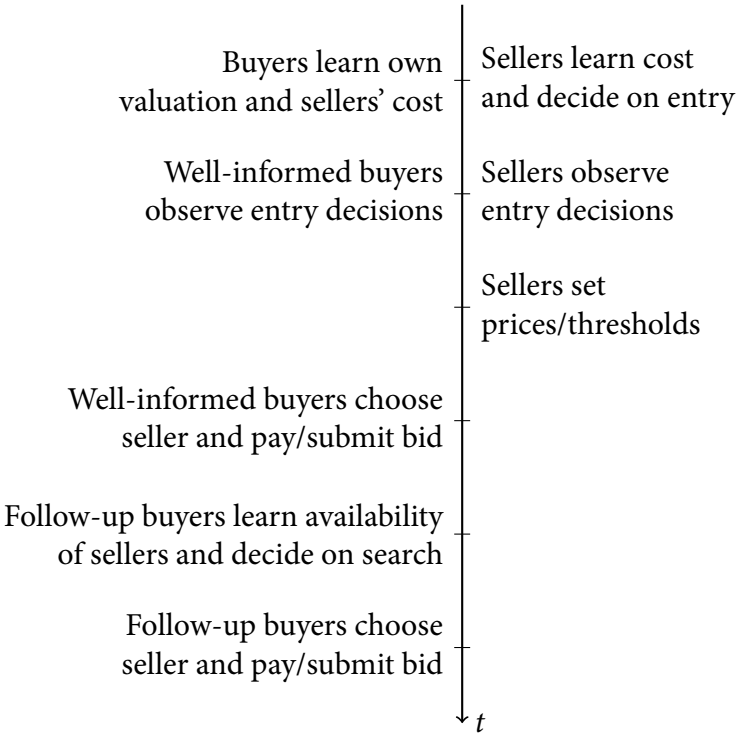


FIGURE 2.5.—Sequence of Events Endogenous Benefits

only one seller, then only this seller is available to follow-up buyers. We summarize this structure in Figure 2.5.[26]

2.5.2 *Results*

Table 2.8 reports the total profits and market shares achieved in PCEB and NCEB. The results are striking. If the PWYW seller chooses to enter, then he is extremely successful in monopolizing the market. On average, his market share is almost 90 percent. In fact, almost all well-informed buyers choose the PWYW seller and follow-up buyers never get to see the posted-price seller. However, this strategy is not very profitable. If costs are low ($c = 5$), the PWYW seller makes a small profit, but if costs increase, PWYW makes losses because buyers do not pay enough voluntarily to

26 Due to space constraints we do not depict the nodes where buyers have the opportunity to purchase after having submitted an unsuccessful bid (given that both sellers are available).

	PCEB				NCEB			
	Profit		Market Share		Profit		Market Share	
	PP	PWYW	PP	PWYW	PP	NYOP	PP	NYOP
$c = 5$	3.4	8.9	11%	89%	22.5	50.5	32%	64%
$c = 10$	5.8	-5.8	11%	86%	21.9	38.6	28%	53%
$c = 20$	7.5	-54.4	25%	75%	18.8	36.8	27%	66%
$c = 30$	0	-97.5	0%	79%	16.4	29.6	23%	36%
$c = 50$	0	-174.0	0%	92%	15.2	21.5	11%	28%
Total	4.2	-9.0	10%	87%	20.4	38.8	27%	54%

TABLE 2.8.—Profits and Market Shares

Notes: Cells show average per period profits and market shares generated by the different seller types in treatments PCEB and NCEB, conditional on entry of both sellers.

cover the seller's costs.[27] Thus, direct promotional benefits alone are not sufficient to make PWYW profitable.

A completely different picture arises in NCEB. If a NYOP seller enters the market his profits are substantial and 70 percent higher than the profits of the posted price seller, but he monopolizes the market less often than a PWYW seller. We observe that a large majority (73 percent) of the well-informed buyers make an offer to the NYOP seller. However, if their bids fail, they turn to the PP seller. Thus, follow-up buyers get to see the PP seller more often. Even if they only see the NYOP seller they often search for an additional seller before making their purchase. The average market share of the NYOP seller is twice as high as the average market share of the PP seller, but significantly lower than the market share of the PWYW seller.

Result 2.5 (Direct Promotional Benefits)**.**

(a) In PCEB, PWYW sellers capture 90% of the follow-up market. However, they fail to make significant profits from this high market share even if costs are low. Thus, word-of-mouth advertising alone generates benefits in terms of market share but not in terms of profits.

27 It has to be noted that very few PWYW sellers entered the market when costs where high. Thus, the high average losses reported in Table 2.8 are based on very few observations.

	PCEB		NCEB	
	PP	PWYW	PP	NYOP
$c = 5$	1.7	2.3	15.4	28.0
$c = 10$	3.7	-3.8	13.2	23.4
$c = 20$	2.5	-44.5	12.1	22.3
$c = 30$	0	-71.5	11.8	12.3
$c = 50$	0	-104.5	5.3	8.5
Total	2.4	-8.0	12.7	21.9

TABLE 2.9.—Endogenous Benefits PWYW & NYOP

Notes: Cells show average per period profits generated by the different seller types on the follow-up markets in treatments PCEB and NCEB, conditional on entry of both sellers.

(b) In contrast, in NCEB there are additional monetary benefits that arise endogenously through word-of-mouth advertising. NYOP sellers attract a twice as large market share on the follow-up market as compared to their PP competitors which translates into substantial profits that are 70 percent higher than the profits of competing PP sellers.

In order to measure the word-of-mouth benefits consider the profits obtained in the follow-up market which are reported in Table 2.9. In PCEB the profits of the PWYW sellers from the follow-up consumers are close to zero or strictly negative. Profits of the competing PP sellers are also close to zero because their market share is very small. However, in NCEB profits on the follow-up market of the NYOP sellers are always positive and almost twice as large as the profits of the competing PP seller. Profits per transaction are very similar across seller types: a PP seller makes 13.5 points on average, while a NYOP seller gets 10.6 points. This suggests that higher profits are not driven by increased profit margins, but by the increased sales volume due to word-of-mouth advertising.

2.6 CONCLUDING REMARKS

Our analysis shows that PWYW and NYOP are effective methods of endogenous price discrimination. Both of these marketing strategies delegate (some) pricing power to buyers, and both strategies avoid setting a refer-

ence price. However, despite these similarities the two pricing strategies work very differently and should be used under different circumstances.

PWYW price discriminates by appealing to social preferences. There is a significant fraction of the population that is willing to voluntarily pay positive prices, which generates some revenues. Prices paid increase with the consumer's valuation, with his prosociality, and with the seller's cost. However, these revenues alone are often not sufficient to cover costs - especially in large anonymous markets in which customers shop only once and have no interest in keeping the firm in business. In such situations, an additional promotional benefit (such as increased press coverage or word-of-mouth recommendations), as well as low marginal costs may be necessary for PWYW to be viable. Because PWYW achieves (almost) full market penetration, it is a very aggressive strategy driving other posted-price sellers out of the market.

NYOP achieves price discrimination by creating strategic uncertainty. Low valuation buyers find NYOP very attractive because it offers a chance to get the good at a lower price. High valuation buyers prefer a posted-price seller, because they do not want to take the risk of not getting the good if their bid is below the seller's secret threshold. Thus, NYOP can be employed to access new customer segments that did not buy the good beforehand while most existing (high valuation) customers still buy at the posted price. Notably, social preferences do not affect how much consumers bid. NYOP has the advantage that the seller can protect himself against losses by setting an appropriate threshold. Therefore, NYOP can also be used profitably if there is no additional benefit and if marginal costs are high. The drawback of NYOP is that it is less successful in penetrating the market and leaves room for additional posted-price sellers. In fact, NYOP is a strategy that relaxes price competition and increases total industry profits.

We find that direct promotional benefits can arise endogenously through a word-of-mouth channel. For PWYW the main benefit is that it monopolizes the market, but it fails in generating significant profits, in particular if costs are high. This suggests that PWYW requires indirect promotional benefits (such as the promotion of complementary goods offered by the same seller) to be viable. The examples of Radiohead and ibis, as discussed in the introduction, are a case in point. NYOP is less successful in terms of market share, but more successful in generating profits. In fact, profits of a NYOP seller are 70 percent higher as profits of the competing PP seller in our experiment.

From a managerial perspective, PWYW is most likely to be successful if capacity constraints are negligible, marginal costs are low and the

seller profits from spillover effects on complementary products. It is a very aggressive strategy that achieves high market penetration but low profits. NYOP, on the other hand, can be employed even if costs are high, and it relaxes price competition. Furthermore, NYOP can be used as a complementary way to sell excess capacity via third-party intermediaries on a permanent basis. It successfully segments the market into high-valuation customers who are more inclined to buy at posted prices and low-valuation customers who would not have bought the good at regular prices. NYOP can therefore be employed parallel to posted prices in order to access new customer segments.

Our study raises several new questions and opens avenues for future research. First, our laboratory analysis is well suited to identify causal effects and to understand the functioning of customer-driven pricing mechanisms. However, it does not tell us much about the magnitude of the observed effects in real markets. Therefore, testing our predictions in the field would be very interesting. Second, we test word-of-mouth as one channel for direct promotional benefits. Further research should explore indirect promotional benefits in more detail. Finally, we ignored some aspects of customer-driven pricing mechanisms, such as joint bidding in NYOP or public minimum/recommended prices, that merit future research.

3

CONTEXT DEPENDENCE IN SPEED DATING

3.1 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 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. 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 & Szeidl, 2013; Bushong et al., 2015).[1] Somewhat disconcertingly, however, these models are based on opposing assumptions about the mapping from utility ranges to decision weights.

In this chapter, I propose a direct empirical test of context dependence using data from a speed-dating experiment previously reported in Fisman et al. (2006), Fisman, Iyengar, Kamenica, and Simonson (2008) and Bhargava and Fisman (2014).[2] I argue that this experiment provides

1 Another class of models is due to Bordalo et al. (2012, 2013). 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 et al. (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),

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 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 & Szeidl, 2013) and relative thinking (Bushong et al., 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 et al. (2015). This effect is more pronounced for women than it is for men, and survives extensive robustness checks.

To understand how individuals make choices is a fundamental question with wide-ranging implications for economic theory and policy.

---

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 3.4.3.

Taking the canonical *homo oeconomicus* as a benchmark for rational behavior, we expect economic agents to have a well-defined and stable preference relation over a known set of alternatives, which in turn guarantees that preference can be revealed through choice. A number of systematic deviations from this definition of rationality have been identified, and they question the validity of *homo oeconomicus* as a positive model. The purpose of this chapter is to provide empirical evidence for a specific type of preference instability, which I will refer to as *context dependence*. In short, an individual suffers from context dependence if she values an alternative differently depending on the consideration set she is facing.

Models of context dependence are useful 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.

As pointed out before, Kőszegi and Szeidl (2013) and Bushong et al. (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 et al. (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, Gerhardt, Riener, Schwerter, and Strang (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 et al. (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, & Camerer, 2012), are tentatively supported by evidence on labor supply and wage expectations (Bracha, Gneezy, & Loewenstein, 2015), and provide novel insights into "first-of-the-month effects" (Huffman & 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 et al. (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 ex-

3 Bushong et al. (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.

pressing 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-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.

3.2 RELATED LITERATURE

Cognitive psychologists have identified several instances in which context influences *perception* (for an overview and classification, see Todorović, 2010). More relevant for the purposes of this chapter, however, is the fact that perceptual biases, caused by the context in which options are presented, impact choice—thus affecting the domain of *preferences*. Early demonstrations of context effects have been documented by Parducci (1965, 1974), Huber, Payne, and Puto (1982) and Huber and Puto (1983). Parducci (1965) develops a theory of perception that remains influential to this day. He postulates two conflicting principles that guide judgment: when presented with a set of stimuli, individuals tend to assign them to categories of equal size by their relation to the range of possible outcomes (for instance, very cold to very warm). At the same time, stimuli are judged by the frequency with which they appear, and hence more frequent stimuli should correspond to larger categories. Whenever stimuli are presented disproportionately to their range, these two principles disagree and decisions are made by compromising between them. Inspired by Parducci (1974), where range-frequency theory is applied to context effects, Huber et al. (1982) and Huber and Puto (1983) show that adding asymmetrically dominated alternatives ("decoys") systematically distorts choice. Simonson and Tversky (1992) build on this literature by developing a unified conceptual framework that can accomodate most of the observed empirical findings.

By now, interest in context dependence and salience has spread from psychology to strands of literatures in business administration and political science, as well as in international communication and economics. Salience as a concept is of course difficult to pin down, and so the precise meaning of the term varies across disciplines.[4] However, it seems to be

4 I will be somewhat liberal in using the terms *salience* and *context dependence* interchangeably. To justify this, the theories outlined below define salience as overweighting

commonly accepted that a salient object is one that "stands out" relative to some other object in a given choice context (likewise for attributes of objects) (cf. Sullins, 1989; Jarvenpaa, 1990; Augoustinos, Walker, & Donaghue, 2014; Kiousis & Wu, 2008; Bordalo et al., 2013).[5]

There are several empirical demonstrations of salience effects in a variety of choice contexts. Jarvenpaa (1990) shows experimentally that it matters whether information is presented as text or in a visual display. She argues that visualizing information triggers perceptual processes that compare physical shapes of items, which may induce decision makers to act on salient features. By contrast, when presenting the same information as numbers, it is much harder to identify and react to dissimilarities.[6] She then exploits the variation in information displays to show that subjects react to features that stand out when presented with the graphic display, but not when information is given as text. In addition, she finds that prior ratings of the importance of attributes are predictive for behavior in the alpha-numeric display, but not in the graphic visualization.

Building on this line of research, Sun, Li, and Bonini (2010) explore whether different *graphical* presentations affect evaluations of options by making dimensions more or less salient. Like Jarvenpaa (1990), they do not vary the informational content of the graphs. But rather than juxtaposing textual and visual displays, their design directly exploits biases in visual perception by manipulating axis scales. To be more precise, they use a hypothetical choice between two scholarships that are characterized by two dimensions, waiting time (in months) and amount of money (in 100 Yuan), where the former is depicted on the horizontal

---

of a specific attribute of an option, which in turn depends on other options that are present in the choice set. In other words, the choice context directly impacts which attributes of an object are considered to be salient, and so it seems warranted to use these two terms interchangeably.

5 Kiousis and Wu (2008) do not explicitly mention that an object needs to "stand out" in order to be considered salient. However, in line with Jarvenpaa (1990) and Bordalo et al. (2013), they define salience in relative terms: "Object salience was defined as the frequency of stories mentioning foreign nations in *New York Times* content [...]." (Kiousis & Wu, 2008, p. 65). The frequency approach immediately implies that when comparing two objects (say, stories mentioning foreign nations and stories mentioning baby names), their salience will be judged in relative terms. For a comprehensive overview of the concept of media salience, see Kiousis (2004).

6 The following quote will further clarify the idea: "A bar representing an information value of 2 is twice as large in appearance as a bar representing an information value of 1. In contrast, a number of 2 versus a number of 1 in an alpha-numeric display possesses fewer physical dissimilarities and does not consume any more display space." (Jarvenpaa, 1990, p. 251).

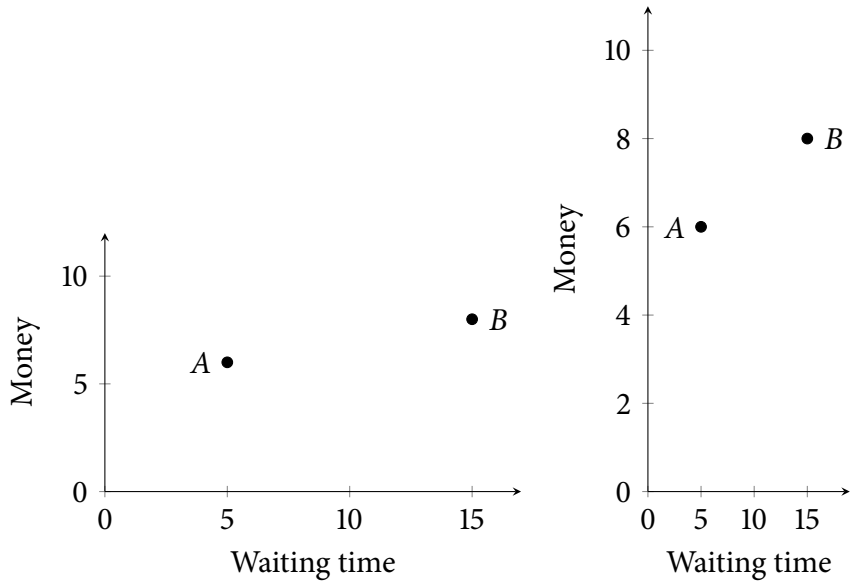


FIGURE 3.1.—Visual Salience

Notes: Adapted from Sun, Li, and Bonini (2010).

and the latter on the vertical axis. The first graph makes waiting time salient by extending the length of the horizontal axis, such that the horizontal distance between the points is much larger than the vertical one. In the second graph, axis lengths are reversed, but the options are equidistant in both dimensions. Figure 3.1 illustrates the design.

Their findings indicate that subjects prefer *A* over *B* when waiting time is salient (i.e., in the left panel of Figure 3.1), while they are indifferent between the two options when the problem is presented as in the right panel of the figure. This is an important finding for two reasons: first, it shows that perceptive errors can affect decision making; second, it implies that preferences can be constructed, at least to some degree. In subsequent experiments, Sun et al. (2010) validate that their main finding is robust to including more than two options and to keeping scale resolution constant across graphs.[7]

An implicit assumption of these findings on visual salience effects is that a dimension becomes more salient the further apart the values of options are. As shown by Sun et al. (2010), the difference in attribute values may not be measured in utility terms, but rather by means of a Euclidean metric. This alludes to a well-established literature on discriminability

7 See Sun, Li, Bonini, and Su (2012) for a more general discussion of graphical framing effects that influence decision making via perception.

(cf. Archer, 1962; Trabasso, 1963; Imai & Garner, 1965). Kahneman (1973) synthesizes their ideas as follows: "Some factors that make a particular cue more salient than others have been identified. Discriminability is such a factor. For example, if ellipses are presented which vary greatly in overall size and only slightly in eccentricity, and both size and shape are relevant, then size rather than shape will dominate behavior in a concept-identification task [...]." (p. 101). Like the previously described experiments on visual salience effects, this definition argues that large differences draw attention, and that heightened perception of a dimension will influence behavior.

3.3 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 et al. (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: $c = (c_1, \ldots, c_K)$.

In line with Kőszegi and Szeidl (2013) and Bushong et al. (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 $\tilde{U}$:

$$\tilde{U}(c, C) = \sum_{k=1}^{K} g_k u_k(c_k), \tag{3.1}$$

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 et al. (2015). To account for the sequential nature of the speed dating experiment, I include previous observations in the consideration set (see Assumption 3.1). That is, all options encountered up to the current partner determine the range of options, and thereby the weight.

Assumption 3.1 (Consideration Sets)**.** *Consideration sets are time dependent and contain all previously encountered options. Formally,* $C^T = \{c^t\}_{t=1}^{T}$.

Assumption 3.2 (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* $\Delta_k(C^T) = \max_{c \in C^T} u_k(c_k) - \min_{c \in C^T} u_k(c_k)$.

Assumption 3.2 states that weights are determined as a function of the difference between the highest and lowest attribute level encountered along a given dimension. 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 et al. (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 et al. (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 3.3 (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 et al. (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 3.4 and the discussion in Section 3.5.1).

Assumption 3.4 (Reference Dependence)**.**

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

$$\tilde{U}^T(c, C^T) = \sum_{k=1}^{K} g_k^T m_k(c_k, r_k), \tag{3.2}$$

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

(b) $\forall c_k$*, 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 3.4 introduces reference dependence into the setup.[8] The aspiration (or reference) level is denoted by r_k, and may thus vary

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

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.[9] Part (b) of Assumption 3.4 is a monotonicity requirement, ensuring that attribute utility functions are increasing, decreasing, or flat over their entire domain.

Together, Assumptions 3.3 and 3.4 imply that

$$m_k(\cdot)\begin{cases} > 0 & \text{if } c_k > r_k, \\ = 0 & \text{if } c_k = r_k, \\ < 0 & \text{if } c_k < r_k. \end{cases} \tag{3.3}$$

Intuitively, expression 3.3 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 & Szeidl, 2013), while the opposite effect is expected for the Bushong et al. (2015) framework to hold true.

Consider again expression (3.2). Keeping the attribute level constant, the marginal effect of an increase in the range on utility is given by

$$\frac{\partial \tilde{U}^T(\cdot)}{\partial \Delta_k(C^T)} = g_k'(\Delta_k(C^T))m_k(c_k, r_k). \tag{3.4}$$

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 et al. (2015) make the exact opposite prediction. To summarize:

9 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 3.5.1.

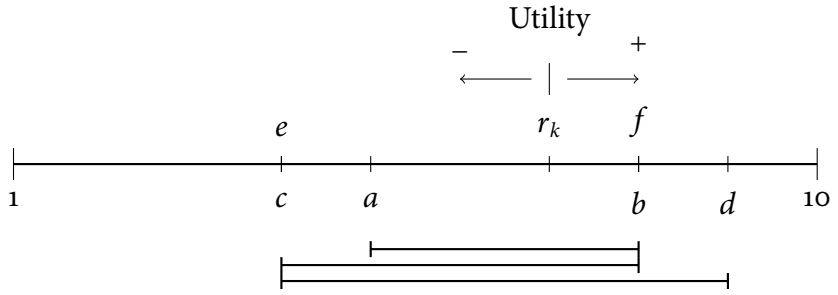


FIGURE 3.2.—Comparative Statics

Prediction 3.1 (Impact of Range on Utility)**.**

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

$$\frac{\partial \tilde{U}}{\partial \Delta_k} = g_k'(\Delta_k(C^T))m_k(c_k, r_k) \begin{cases} > 0 & \text{if } m_k(\cdot) > 0, \\ < 0 & \text{if } m_k(\cdot) < 0 \end{cases} \tag{3.5}$$

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

$$\frac{\partial \tilde{U}}{\partial \Delta_k} = g_k'(\Delta_k(C^T))m_k(c_k, r_k) \begin{cases} < 0 & \text{if } m_k(\cdot) > 0, \\ > 0 & \text{if } m_k(\cdot) < 0. \end{cases} \tag{3.6}$$

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 3.4.4.

3.4 EMPIRICAL ANALYSIS

3.4.1 *Institutional Background*

Speed dating was invented by Rabbi Yaacov Deyo in 1998 and has spread rapidly since, with only minor variations to the protocol.[10] 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 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.

3.4.2 *Methods and Data*

The data used in this chapter 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.

3.4.2.1 *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 for the duration of 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

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

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 3.1.—Summary of the Speed Dating Sessions

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.

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.[11]

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 3.1 gives an overview of all sessions that are included in the data analysis.

11 Including all sessions in the analysis leaves the main results qualitatively unchanged (see Appendix C.1.1).

3.4.2.2 *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.[12] For the purposes of this study, 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 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.

3.4.3 *Preliminary Considerations*

SPECIFICATION OF THE ATTRIBUTE RANGE I assume throughout the main text 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 3.5.4 and Appendix C.1.4).

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

	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 3.2.—Importance Ratings

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 C.1.2). Since specifying the attribute range at the first date requires an additional assumption, 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 3.2, 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).

TIME TRENDS Another concern that may influence the estimated effects comes from the sequential nature of the speed dating experience. As shown in Figure 3.3, 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.[13]

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 indeed 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.[14]

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 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

13 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.

14 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).

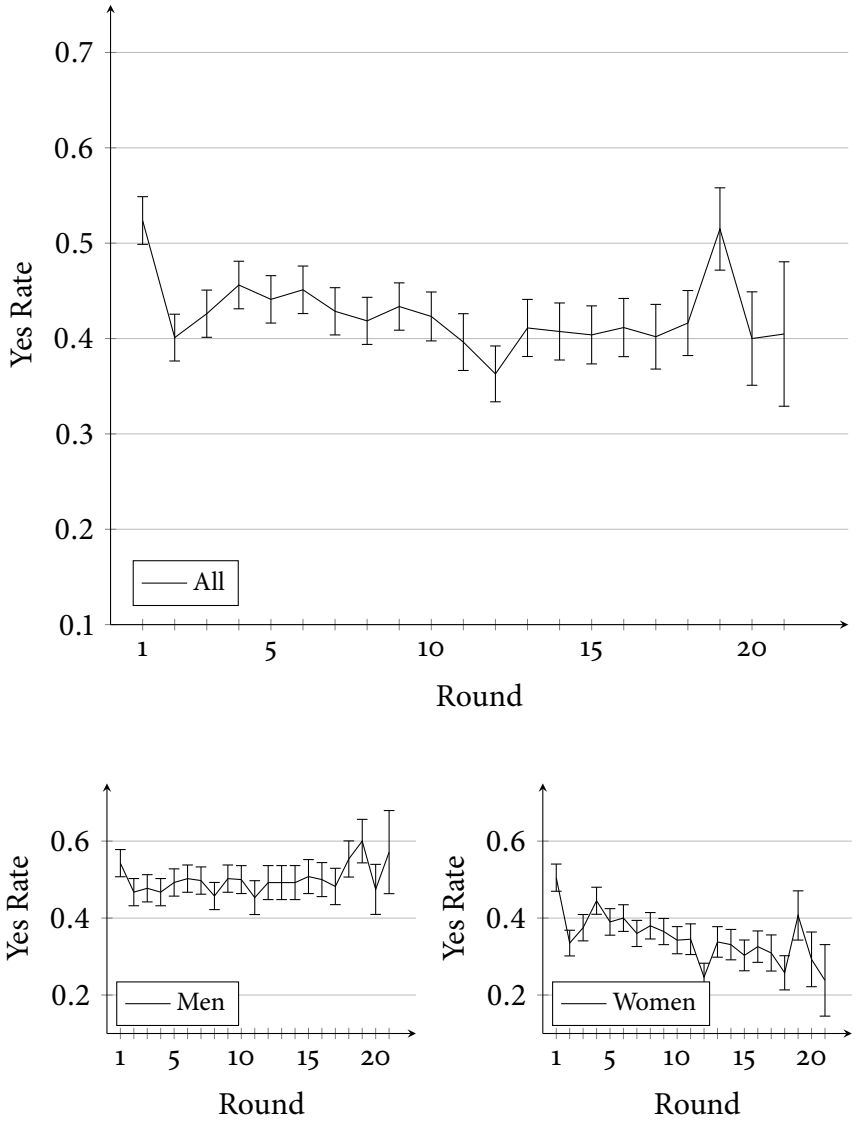


FIGURE 3.3.—Yes Rates

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.

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 C.1.3.

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.

This sequence can then be tested for stationarity using an augmented Dickey-Fuller test. Doing so for each attribute and for the whole sample as well as men and women separately reveals that the null hypothesis of stationarity cannot be rejected for 16 out of 18 cases.[15] Failure to reject stationarity for 16 of the 18 time series implies that over- and underrating follows a random walk, hence there is no systematic recalibration over time.

There is a significant positive trend for seven of these time series, meaning that subjects in earlier rounds are underrated, while those in later rounds tend to be overrated. 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.

15 There are 6 attributes (attractiveness, sincerity, intelligence, fun, ambition and shared interests) and three subsamples (all, men, women) for a total of 18 augmented Dickey-Fuller tests. Stationarity is rejected for attribute 'fun' when using the subsample consisting only of women (MacKinnon approximate p-value < 0.05) and for attribute 'shared interests' when using the subsample consisting only of men (MacKinnon approximate p-value < 0.05).

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 & DiNubile, 1976).[16] Considering the speed dating experiment, a person's attractiveness may be judged less favorably when preceded by a very attractive person. However, 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.

3.4.4 *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.

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

As a largely realistic first pass, I consider the evaluator's own level along an attribute as a natural threshold (Figures 3.4 and 3.5, 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

$$\begin{aligned}\text{Decision}_{ijt} = \alpha_i + \gamma_t &+ \sum_{k\in K} \beta_{k1} \cdot \text{Range}_{ikt} \\ &+ \sum_{k\in K} \beta_{k2} \cdot (\text{Rating}_{ijkt} > \text{Threshold}_{ik}) \cdot \text{Range}_{ikt} \\ &+ \sum_{k\in K^+} \beta_{k3} \cdot \text{Rating}_{ijkt} + \sum_{k\in K} \beta_{k4} \cdot \text{Rating}_{ijk,t-1} + \varepsilon_{ij},\end{aligned} \tag{3.7}$$

where $K = \{\text{Attractiveness, Fun}\}$, $K^+ = \{\text{Attractiveness, Sincerity, Intelligence, Fun, Ambition, Shared Interests}\}$, 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.

3.5 RESULTS

3.5.1 *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 et al. (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 3.4 and 3.5 presents results of five separate regressions of the decision to date on each attribute, including time and individual fixed effects. The solid black line plots linear predictions for every possible level of the attribute, and corresponding 95% confidence intervals are shown as grey areas.

Horizontal red dashed 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 dotted line); and the average consensus rating of subjects, explained in more detail below (blue dash-dotted line).

Figures 3.4 and 3.5 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 3.4.4.[17] These observations are summarized in the following result:

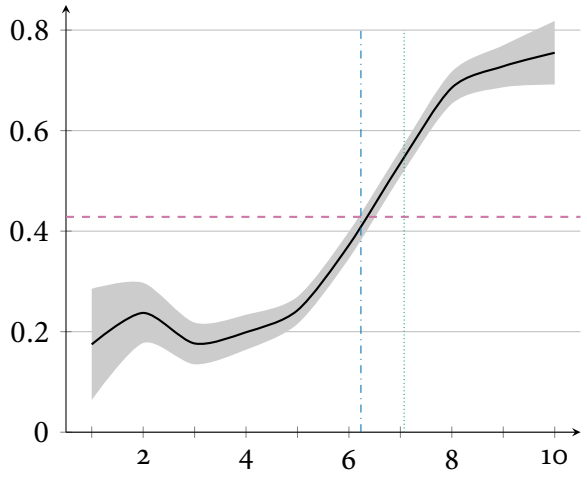
Result 3.1 (Attribute Utility Functions & Thresholds)**.**

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

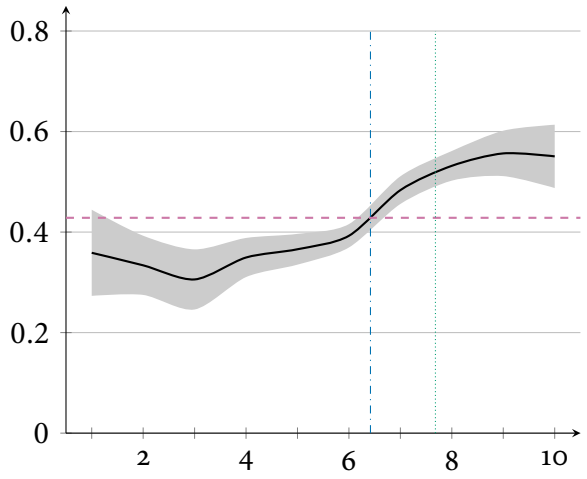
(b) Utility increases monotonically in attractiveness and fun almost everywhere. Attribute levels above the average consensus rating are perceived as 'good', while those below are perceived as 'bad'.

Result 3.1 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

17 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 chapter and should therefore be regarded as speculative.



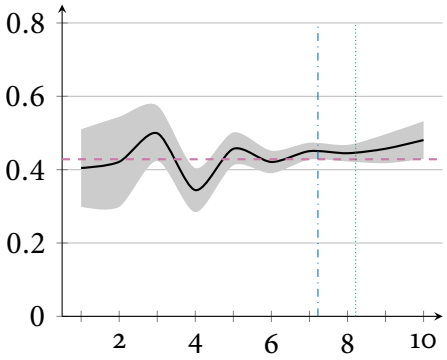
(A) Attractiveness



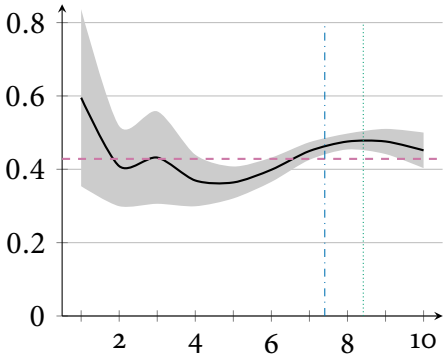
(B) Fun

FIGURE 3.4.—Marginal Effects of Attribute Levels on Dating Decision (Attractiveness & Fun)

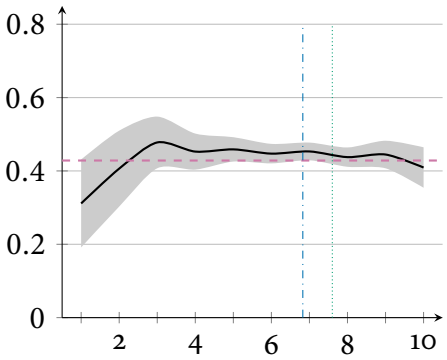
Notes: Vertical axes display average predicted values of the dependent variable (decision to date) for different attribute values. Plotted values are derived from regressing the decision to date on all five attributes, including time and individual fixed effects. Marginal effects are obtained post estimation 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 3.5.—Marginal Effects of Attribute Levels on Dating Decision (Sincerity, Intelligence, & Ambition)

Notes: The same notes as for Figure 3.4 apply.

implication, the identification strategy proposed in Section 3.4.4 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.

3.5.2 *Context Dependence*

Tables 3.3 and 3.4 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.

All specifications rely on the ratings made by participants at the end of each date. In Table 3.3, 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 3.4 differs from Table 3.3 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 3.4 and 3.5), the following analysis will focus on the results in Table 3.4.

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

	Own Rating		
	(1) All	(2) Male	(3) 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 3.3.—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 3.4.—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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

effect is negative. However, the effects are jointly insignificant and thus indistinguishable from zero.[18]

To probe further into these results, columns (2) and (3) split the sample by gender.[19] 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 et al. (2015). The joint effect is significantly negative in this case ($F(1,199) = 6.55$, $p < 0.05$). It is insignificant for men.[20]

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 3.4). In combination with the smaller effect size, this makes it more difficult to pick up the effect.

Result 3.2 (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.

18 Results are qualitatively unchanged when controlling for a linear time trend instead. The joint effects on attractiveness in columns (1) and (3) of Table 3.3 and column (3) of Table 3.4 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 3.3, insignificant in specification (3) of Table 3.3, and significantly negative at the 5% level in specification (3) of Table 3.4. Magnitudes of coefficients are very similar.

19 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 study is the first to find a gender gap in context dependence, speaking to the domain-specificity of such differences.

20 Note that these results are not distorted by differential utility functions of male and female participants. Figure C.1 in Appendix C.2 shows that both men and women exhibit utility functions that are 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.

(b) Results for both attractiveness and fun are qualitatively consistent with Bushong et al. (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 3.2 supplements and substantiates the graphical analysis presented in Figure 3.4. 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 et al. (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 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.

3.5.3 *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 3.4 and 3.5 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

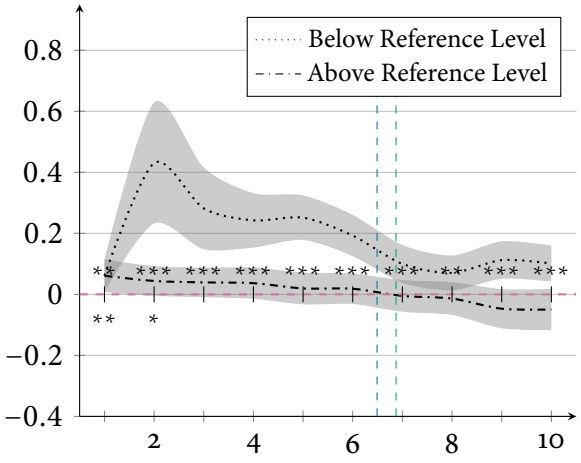
fact that the estimated utility function passes through the intersection point and is monotonic suggests that the consensus rating serves as a natural reference level for relative utility judgments.

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.

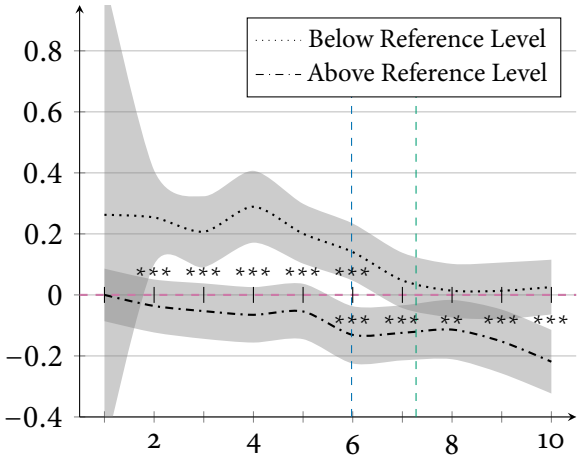
Figure 3.6 plots the results for attractiveness. I run regressions of the same type as in Section 3.5.2, except for one crucial difference: I hold the interaction reference level for fun constant at the consensus rating, while varying the reference level for attractiveness from 1 to 10. The dotted line 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). Stars 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 stars below the red line and coefficients of the above reference level attribute range.

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 et al. (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



(A) Men



(B) Women

FIGURE 3.6.—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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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 3.6b 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.[21]

Result 3.3. *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 is not satisfied: 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 C.3).*

To put this in plain words, Panel 3.6b 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.[22] 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

21 Performing the same type of analysis on the combined sample, Panel C.2a in Appendix C.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 3.4), this finding is expected.

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

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.[23]

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.

3.5.4 *Consideration Set Composition*

Recall that by Assumption 3.1, 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 3.5 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 3.5 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 ($F(1, 399) = 3.02$, $p < 0.1$; $F(1, 399) = 2.72$, $p < 0.1$).

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

23 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 3.5.—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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the evaluator is insignificant for male evaluators, but highly significant ($p < 0.01$) and negative for female evaluators (cf. Appendix C.1.4).

Result 3.4. *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 C.1.4).

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.[24] The robustness of these results should therefore be jugded relative to the validity of the approach upon which they are based.

3.6 CONCLUDING REMARKS

This chapter 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 from 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 et al. (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

24 Masatlioglu et al. (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 chapter, I therefore rely on a simplified framework.

same utility difference receives 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 study 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 & Kurtz, 2013). The findings of this study suggest ways in which such targets can be reached more effectively.[25]

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 study 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

25 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).

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 chapter therefore lies 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 and directionally similar for female and male evaluators, coefficients on 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 et al., 2012; Manzini & 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.

Part III

APPENDIX

A

APPENDIX TO CHAPTER 1

A.1 INSTRUCTIONS

Thank you for your participation!

For your participation you will receive a base amount of 4 Euros. You will be able to earn additional money in the following experiment. You will receive your money at the end of the experiment in an envelope. Neither the other participants of this study nor the experimenters will be able to see how much money you have received.

Please note: during the entire duration of the experiment, **communication between participants is not allowed**. If you have a question, please raise your hand. Your question will then be answered privately.

Please note: **all statements made in these instructions are true**. This is true for all experiments conducted at the Bamberg Laboratory for Experimental Research, and also for this experiment. **In particular, all consequences that are described in the instructions will be carried out exactly as described**. If you wish, you may verify the correctness of all statements made in these instructions after the experiment.

In this experiment there are **two options**. Depending on which option you choose, you will be able to earn different amounts of money. In addition, there will be different consequences for a laying hen, depending on your chosen option.

Details Regarding the Laying Hen

In this study, a laying hen is entrusted to your care. It is a young and healthy hen. The hen is now old enough (18 weeks) to be put into a coop with other laying hens on a Franconian farm. In this regard, there are two options for the laying hen among which you may choose. The hen will live for another approximately 13 months from the time it is put into the coop.

How do the options differ?

Conventional Option: If you choose option conventional, you will receive an **additional amount of money** at the end of the experiment and the hen will be put into a **conventional** farming environment.

FIGURE A.1.—Pictures of a Chick and a Grown Hen

Organic Option: If you choose option organic, you will **not receive any additional amount of money** at the end of the experiment and the hen will be put into an **organic** farming environment.

Details Regarding the Living Conditions

Conventional Option: In the conventional farming environment, each hen has an available average space of 0.17 square meters. This amounts to an area of 41 cm × 41 cm. A pawing area covers approximately one third of the coop. In the remaining two thirds there is a fecal pit with perches above it. Your hen does not have the possibility to go outside. The feed is from conventional agriculture, antibiotics are allowed.

Organic Option: In the organic farming environment, your hen has 4 square meters of grassy outdoor runs as well as a canopied outdoor area with air-conditioning at its disposition, in addition to the area in the coop. Thus, it can walk around on the grass together with its fellows, take a sand bath, pick the ground and experience the fresh air. In addition, it receives organic feed that must satisfy strict standards. Antibiotics may not be fed. The conditions of the "Naturland"-certificate are in many ways stricter than the regulations of the EU ecological certificate: for example, the criteria for feed are stricter, the maximum amount of hens per coop is lower and the animals have the opportunity for free-range activity even in bad weather periods.

The following table gives an overview of the different living conditions and provides you with additional information:

	Conventional	Organic
grassy outdoor runs	no	~ 4 m^2 per hen
canopied outdoor area	no	available year-round, possibility to sand- and dustbathe
size of coop	up to 6000 hens	up to 3000 hens
feed	antibiotics allowed	no antibiotics allowed

TABLE A.1.—Comparison of Living Conditions

Details Regarding the Payment

In a few moments, you will be presented with a list of choices. On this list, different amounts of money for choosing the option conventional will be displayed: it starts out with 50 Cents and increases in 50-Cent-steps to 25 Euros. For each of these amounts you can choose between conventional and organic.

Your choice for each of these amounts is important. The computer will randomly select one of the amounts at the end of the experiment. **Your choice for this amount of money will be executed with all consequences.** Should the computer choose 3 Euros and should you have checked the box for conventional at 3 Euros, the laying hen will be put into the conventional farming environment and you will receive an additional 3 Euros at the end of the experiment. Should the computer choose 22.50 Euros and should you have checked the box for conventional at 22.50 Euros, the laying hen will be put into the conventional farming environment and you will receive an additional 22.50 Euros at the end of the experiment. Should the computer choose 15 Euros and should you have checked the box for organic at 15 Euros, the laying hen will be put into the organic farming environment and you will not receive any additional money at the end of the experiment.

Summary

You are provided with a list of decisions. This list contains different amounts of money for the choice of the conventional option. For each amount of money, you choose between the conventional and the organic option. The computer randomly selects one of these amounts. If you

have chosen the conventional option at this amount, you will receive the additional amount of money at the end of the experiment and the hen will be kept in the conventional farming environment for the rest of its life. If you have chosen the organic option at this amount, you will not receive any additional money, and the hen will be kept in an organic farming environment (Naturland) for the rest of its life.

A.2 PRIMING EFFECTS

Before eliciting subjects' willingness to pay for sustainability, we randomly exposed them to three different texts.[1] In the control treatment (C), subjects received two texts, one taken from German magazine FOCUS online and one by the renowned physicist and biochemist Francis Crick. These texts dealt with the concept of consciousness and its relevance for medical applications and psychology.

In the second treatment (AFW), subjects were given two texts that argued against the existence of free will. One of the texts was taken from FOCUS online and explained, in a rather summary manner, the basics of the Libet experiment in the 1980s (Libet, 1985) and a recent contribution from a French group of researchers (Desmurget et al., 2009), claiming that people do not have free will. The other text, again by Francis Crick, argued strongly in favor of a deterministic worldview, in which intentional actions and hence free will are mere illusions.

In the third treatment, we introduced an anti-religion prime (AR), consisting of texts by atheist authors Richard Dawkins and Christopher Hitchens (Dawkins, 2009; Hitchens, 2008). These texts were primarily concerned with outlining the wars and crimes that have been committed in the name of religion, as well as arguing against God's existence.

We expected that evoking anti-free will and anti-religious concepts would alter subjects' behavior on the subsequent animal welfare task towards caring less for the living conditions of a hen.

Contrary to our expectations, priming subjects with concepts of determinism and anti-religious thought did not have an effect on our dependent or independent measures of interest. (See Table A.2 for an overview.)

We therefore decided to retain the observations of all treatments (C, AFW and AR) for analyzing the association of ethical behavior towards humans and animal welfare.

1 All texts are available upon request.

Variable	*p*-Value
Switch Point	.810
Openness	.809
Agreeableness	.474
Machiavellianism (Tactics)	.991
Machiavellianism (Cynicism)	.527
Feminism Score	.061
Masculinism Score	.647
Age	.067
Gender	.417
Vegetarian or Vegan	.044
Religiosity Dummy	.923

TABLE A.2.—Differences Between Priming Treatments

Notes: Table reports *p*-values from Kruskal-Wallis tests under the null hypothesis that there are no differences in switch points, demographics, and personality characteristics.

However, note that there are cross-treatment differences in demographic variables like age and gender. This suggests that our sample is somewhat unbalanced across treatments in terms of observables. To address this confound, we present results based on only the control sample in the next section.

A.3 ROBUSTNESS

Table A.3 presents analyses performed on the control sample only, i.e. those subjects that were given neutral text treatments.

Results are qualitatively similar. The effects of openness and Machiavellianism (Tactics) are larger in magnitude and more precisely estimated. Contrary to the findings in the main text, being a vegetarian or vegan is not predictive for a higher switch point when using the control sample.

These results reinforce the point that demographic criteria seem to have little predictive power for the decision about living conditions of animals. On the other hand, variation in demographics may be larger than that in psychographic criteria, and therefore a larger sample size would have been needed to detect statistically significant effects.

	(1)	(2)
	Excluding Multiple Switchers	
Openness	0.957***	0.906***
	(0.315)	(0.331)
Agreeableness	0.0390	-0.181
	(0.485)	(0.516)
Machiavelli (Tactics)	-0.980**	-1.038**
	(0.416)	(0.446)
Machiavelli (Cynicism)	0.0306	-0.233
	(0.613)	(0.614)
Locus of Control	-0.296	-0.104
	(0.277)	(0.275)
Feminism Score	0.0381	0.138
	(0.235)	(0.261)
Masculinism Score	0.0251	-0.189
	(0.229)	(0.223)
Age		0.457
		(0.703)
Gender		-0.106
		(4.765)
Vegetarian or Vegan		6.969
		(7.053)
Religiosity Dummy		-4.454
		(4.740)
Constant	31.19	30.45
	(39.44)	(42.63)
N	89	86
N (uncensored)	42	40
N (left-censored)	14	13
N (right-censored)	33	33
Pseudo R^2	0.029	0.041
Log Likelihood	-220.7	-208.8

TABLE A.3.—Control Sample

Notes: Tobit regressions with switching point as dependent variable. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B

APPENDIX TO CHAPTER 2

B.1 PROOFS

Proof of Lemma 1. Because the threshold is set secretly it cannot affect the behavior of the buyers nor of a competing seller. Note that the threshold has to be nonnegative. Suppose that the seller sets a threshold $\hat{t} > t^*$. Buyers offering a price $p \geq \hat{t}$ receive the good and pay the offered price under both $\hat{t}$ and t^*. Buyers offering a price $p < t^*$ do not receive the good and do not pay anything under both thresholds. Thus, in these cases the threshold does not make a difference. However, if a buyer offers a price p such that $t^* \leq p < \hat{t}$, then the buyer does not get the good under threshold $\hat{t}$ and the seller makes zero profit, while the buyer gets the good under threshold t^* and the seller makes a positive profit of $p - c + b > 0$ from this customer. Thus, the seller is better off with threshold t^* than with threshold $\hat{t} > t^*$. Similarly, suppose that the seller sets a threshold $\hat{t} < t^*$. Again, if $p < \hat{t}$ or if $p \geq t^*$ it does not make a difference whether the threshold is $\hat{t}$ or t^*. However, if $\hat{t} \leq p < t^*$ then the buyer gets the good under threshold $\hat{t}$ yielding a negative profit of $p - c + b < 0$ for the seller, while the buyer does not get the good and the seller's profit is zero under threshold t^*. Thus, setting $t^* = \max\{c - b, 0\}$ is indeed a (weakly) dominant strategy. □

Proof of Lemma 2. Two cases have to be distinguished. (1) If both sellers are posted-price sellers (which may happen in the flexible competition treatment) the unique (Bertrand) equilibrium of the this game is that both posted-price sellers charge $p^P = c$. (2) If a posted-price seller faces a NYOP seller the NYOP seller sets the threshold $t^* = \max\{c - b, 0\}$ and all buyers buy from him if $p^P > c$. In this case the posted-price seller makes a profit of zero. If he sets $p^P = c$ some buyers may buy from him, but his profit is again zero. Charging $p^P < c$ can only yield losses and is dominated by $p^P = c$. □

Proof of Lemma 3. If the buyer makes an offer to the NYOP seller he should either offer $p = c$ which gets him the good with certainty or $p = \max\{c - 40, 0\}$ which is successful with probability 0.5. The safe bid is optimal if $v_i - c \geq \frac{1}{2}[v_i - max\{c - 40, 0\}]$, which is equivalent to $v_i \geq \min\{2c, c + 40\}$. □

B.2 INSTRUCTIONS

GENERAL INSTRUCTIONS

Welcome to the Experiment and Thank You for Participating!

From now on, please do not talk to any of the other participants of the experiment.

GENERAL INFORMATION

The purpose of this experiment is to study economic decision behavior. By participating, you can earn money. Following the experiment, you will be paid out in cash.

During the experiment, you and the other participants will be asked to make decisions. Both your own decisions and those of the other participants determine your payoff according to the rules explained below.

The whole experiment will take about two hours. In the beginning you will receive detailed instructions. If you have any questions after the instructions or during the experiment, please raise your hand. One of the supervisors of the experiment will come to your cubicle and privately answer your questions. The question will be repeated and answered publicly if it is relevant for all participants. You will have to answer some control questions after the instructions before we can start the experiment.

For linguistic convenience we use male terms throughout.

Payment

We will talk about points and not Euro in the experiment. These points will be converted into Euro at the end of the experiment. The exchange rate is:

120 points = 1 Euro

In addition, you will receive an initial endowment of 480 points (4 Euro) at the beginning of the experiment. There will be an additional compensation for filling out a questionnaire at the end of the experiment.

Anonymity

Your decisions and payoffs are anonymous: Neither will you receive any information concerning the decisions or payoffs of the other participants, nor will the other participants receive any information concerning your own decisions or payoffs. We will analyze the collected data anonymously and will never link names to data from the experiment. At the end of the experiment you will have to sign a receipt stating that you received your payment. This receipt is used for accounting purposes only.

Miscellaneous

There will be a pen on your desk. Please leave this pen on the table after the experiment.

INSTRUCTIONS FOR FIRST SET OF EXPERIMENTS

The following instructions were presented to subjects in the NCFix treatment. All other instructions of the first set of experiments were adapted accordingly and are available from the authors upon request. Instructions were succeeded by a set of control questions (not shown here) that were checked by the experimenters after the subjects had answered them in private.

THE EXPERIMENT

Roles

There are two roles in the experiment, which we will refer to as buyer and seller in the following. Your role will be assigned to you randomly and will stay the same for the duration of the experiment.

Decisions and Procedures

The experiment consists of 20 periods. In all periods the same kind of decisions will have to be made.

At the beginning of each period, two sellers will be matched with six randomly selected buyers. Each buyer is assigned to exactly two sellers: one of them will be a posted-price seller and the other will be a Name Your Own Price seller. Buyers can purchase exactly one unit

of the offered good in each period if at least one seller has entered the market. Buyers can thus decide from which seller he wants to buy or if he does not want to buy at all.

How do the two sellers differ?

- If the **posted-price seller** offers his good, he sells it through the **posted price mechanism**. This means that the seller chooses a price for which buyers can purchase the good. If a buyer decides to buy the good, he pays the posted price. If he does not buy the good, there will be no trade and both the buyer and the seller will receive a payoff of **0 points**.

- If the **Name Your Own Price seller** offers his good, he sells it through the **Name Your Own Price mechanism**. This means that, in a first stage, the seller determines a price threshold that is known only to him. Each buyer will then decide on his own which price he wants to offer for the good. The seller has to deliver the good only if the buyer has offered a price greater than or equal to the price threshold (which is still unknown to the buyer). If the price is greater than or equal to the price threshold, the buyer will pay exactly the amount he offered; if the price is lower than the price threshold, there will be no trade and both buyer and seller will receive a payoff of **0 points**.

Each period consists of 5 stages:

1. In the first stage of each period, sellers are told their costs for producing a unit. These costs are the same for both sellers. Moreover, the Name Your Own Price seller learns about the benefit that he will receive every time he sells a good. Only the Name Your Own Price seller can receive this benefit; the posted-price seller neither knows about the size of the benefit nor will he receive it.

2. In the second stage of each period, sellers will independently decide whether to offer the good or not. If a seller does not enter the market, he will receive a payoff of **0 points**; in that case, buyers still have the chance to buy the good from the remaining seller. If none of the sellers enter the market, both sellers and their assigned buyers will get a payoff of **0 points** and the period ends.

3. In the third stage of each period, sellers are informed whether the other seller has entered the market or not.

4. If the posted-price seller has entered the market, he determines the price at which the good is available to buyers in stage four of the period. If the Name Your Own Price seller has entered the market, he must specify a price *threshold*. Only offers that exceed this threshold will lead to a transaction.

5. In the fifth stage, buyers learn their valuation for the good. This translates to the payment a buyer will receive at the end of the experiment if he has bought the good. In addition, buyers will be informed about sellers' costs of production for the good. Each buyer now decides whether to buy the good and if he wants to buy it, where to buy it. If he decides to buy from the Name Your Own Price seller, he also has to make an offer for the good.

After each period, two sellers (one posted-price seller and one Name Your Own Price seller) will be assigned anew to six randomly selected buyers.

Detailed Procedures

Each period proceeds as follows:

1. Sellers are informed about their costs of production for each sold unit of the good. Costs can be **10**, **30** or **50 points** and are randomly drawn in each period. Moreover, the Name Your Own Price seller receives a benefit for each unit sold. The Name Your Own Price seller learns whether the benefit is **0** or **40 points**.

2. Sellers decide independently from each other whether to offer the good in this period or not. If a seller does not enter the market, he receives a payoff of **0 points**. In this case, buyers have the opportunity to purchase the good from the remaining seller. If none of the sellers enter the market, both of them and their assigned buyers get a payoff of **0 points** and the period ends.

3. Sellers are informed whether the respective other seller has entered the market or not.

4. If the posted-price seller has entered the market, he must now determine a price. This price is identical for all buyers on the market and can vary between 0 and 200 points.

If the Name Your Own Price seller has entered the market, he must determine the threshold price. This threshold price is also identical for all buyers on the market. All integers between 0 to 200 points are valid thresholds.

5. Buyers are informed about their valuation for the good. This valuation can be **10**, **25**, **40**, **60**, **120** or **200** points and is drawn randomly in each period. It is possible that multiple buyers have the same valuation. In addition, buyers are informed about the unit costs of sellers.

   Subsequently, all buyers decide whether to buy the good and if yes, where to buy it. If the buyer chooses the Name Your Own Price seller, he also needs to decide which price he would like to offer for the good. The mount of this offer can be freely chosen. Each amount between 0 points and the own valuation in points is valid.

6. Buyers learn their income that they have earned in the current period. Buyers who opted for the Name Your Own Price seller are informed whether their offer was successful or not.

   posted-price sellers learn how many buyers bought at the posted price, how many buyers submitted an offer to the Name Your Own Price seller and how much they earned.

   Name Your Own Price sellers learn the prices offers submitted by each of their buyers and whether these offers were successful. In addition, Name Your Own Price sellers are informed about the number of buyers who decided to buy from the posted-price seller and learn their payoff in the current period.

End of a Period

At the end of each period, sellers and buyers are separated and two sellers (one posted-price seller and one Name Your Own Price seller) are assigned to six new, randomly selected, buyers. Hence, a market always consists of one posted-price seller, one Name Your Own Price seller and six buyers. After 20 periods the experiment ends.

CALCULATION OF INCOME AT THE END OF EACH PERIOD

If none of the sellers have entered the market in a given period, both sellers and the respective six buyers in the market receive **0 points**.

BUYERS

Buyers can either buy from the posted-price seller, or from the Name Your Own Price seller, or they can choose not to buy at all.

Transacting with the Posted-Price Seller

If a buyer purchases from the posted-price seller, his income equals his valuation V minus the posted price P_P.

$$\text{Income} = \text{Valuation } (V) - \text{posted price } (P_P)$$

Transacting with the Name Your Own Price Seller

If a buyer purchases from the Name Your Own Price seller and if his offer was successful, his income equals his valuation (V) minus the offered price (P_O).

$$\text{Income} = \text{Valuation } (V) - \text{Offered Price } (P_O)$$

No Transaction

If the buyer does not trade in this period, which can happen when

- None of the sellers have entered the market, or
- The buyer did not buy the good either from the posted-price seller or from the Name Your Own Price seller, or
- The price offer of the buyer was not greater than or equal to the price threshold of the Name Your Own Price seller and was therefore not successful,

the income of the buyer equals **0 points**.

POSTED-PRICE SELLER

The income of the posted-price seller equals his revenues minus his costs.

- His revenue is calculated by multiplying the posted price he determined with the number of goods sold.

- His costs are calculated by multiplying the unit cost (C) with the number of goods sold.

Thus, if the seller has entered the market and three buyers have purchased from him:

$$\begin{aligned}\text{Income} &= \text{Sum of posted prices } (3 \cdot P_P) \\ &\quad - \text{Sum of Unit Costs } (3 \cdot C)\end{aligned}$$

If the posted-price seller has not entered the market, his revenue in this particular period equals **0 points**.

If a buyer does not buy the good from him, there will be no unit costs for this buyer and the seller will not receive any money from this buyer.

NAME YOUR OWN PRICE SELLER

The income of the Name Your Own Price seller is his revenue plus the benefit and minus his costs.

- His revenue is the sum of the offered prices that are greater than or equal to his price threshold, plus the benefit (B) multiplied with the number of goods sold.
- His costs are the unit costs (C) multiplied with the number of goods sold.

Thus, if the seller has entered the market and the price offers of three buyers are successful:

$$\begin{aligned}\text{Income} &= \text{Sum of Offered Prices} \\ &\quad (P_O^1 + P_O^2 + P_O^3) \\ &\quad + \text{Sum of Unit Benefits } (3 \cdot B) \\ &\quad - \text{Sum of Unit Costs } (3 \cdot C)\end{aligned}$$

If the Name Your Own Price seller has not entered the market, his revenue for this particular period equals **0 points**.

If a buyer's offered price is below the price threshold, there are no unit costs and no benefits and he will not receive a price from this buyer.

If a seller or a buyer incurs a loss in any given period, they have to use their previous profits or their initial endowment, respectively, to compensate for this loss.

Do you have any questions up to this point?

INSTRUCTIONS FOR SECOND SET OF EXPERIMENTS

The following instructions were presented to subjects in the NCEB treatment. The corresponding instructions for the PCEB treatment are available upon request. Instructions were succeeded by a set of control questions (not shown here) that were checked by the experimenters after the subjects had answered them in private.

THE EXPERIMENT

Roles

There are three roles in the experiment, which we will refer to as well-informed buyer, follow-up buyer and seller in the following. Your role will be assigned to you randomly and will stay the same for the duration of the experiment. On the first screen you will be informed to which role you are assigned.

Decisions and Procedures

The experiment consists of 20 periods. In all periods the same kind of decisions will have to be made.

At the beginning of each period, two sellers will be matched with six randomly selected buyers (two well-informed buyers, four follow-up buyers). First, every seller decides if he wants to sell on the market. Buyers can purchase a maximum of one unit of the offered good.

How do the two sellers differ?

- The **posted-price seller** chooses a price for which buyers can purchase the good.
- The **Name Your Own Price seller** determines a price threshold that is known only to him. Each buyer will then decide on his own which price he wants to offer for the good. If the price is greater than or equal to the (unknown) price threshold, the buyer will pay exactly the price he offered and gets the good. If the price is lower than the price threshold, there will be no trade.

How do the two buyers differ?

- The two well-informed buyers buy first and may choose from which the seller they would like to buy (if both sellers have decided to sell).

- Then, the four follow-up buyers make their purchasing decisions. They are only able to purchase from a seller if the seller has decided to sell on the market and if one of the two well-informed buyers has bought from that seller or submitted a price offer.
- Both well-informed and follow-up buyers can purchase the good from the posted-price seller if the price offer they submitted to the Name Your Own Price seller was not successful. For well-informed buyers this possibility exists if the posted-price seller has entered the market. For follow-up buyers this possibility exists only if the posted-price seller is available, that is
  - if he has entered the market and at least one well-informed buyer has bought from him or
  - if he has entered the market, no well-informed buyer has bought from him, but the follow-up buyer paid search costs of **10 points**.

Each period consists of eight stages:

1. All participants are told the costs that accrue to sellers for every unit sold. These costs are the same for both sellers. Costs can be **5, 10, 20, 30** or **50 points** and are randomly drawn in each period.

   Buyers are informed about their valuation for the good. This valuation can be **10, 25, 40, 60, 120** or **200 points** and is drawn randomly in each period for every buyer. It is possible that multiple buyers have the same valuation. Sellers are not informed about these valuations.

2. Every seller decides whether to offer the good or not. If a seller does not enter the market, he receives a payoff of **0 points**.

3. Sellers are informed whether the respective other seller has entered the market or not. If the posted-price seller has entered the market, he must now determine a price. If the Name Your Own Price seller has entered the market, he must determine the threshold price. Only offers that exceed this threshold will lead to a transaction. The posted and threshold prices are identical for **all well-informed and follow-up buyers** in this period.

4. The two well-informed buyers learn which sellers entered the market and decide whether to buy the good and if yes, where to buy it. If a buyer chooses the Name Your Own Price seller, he also needs to decide which price he would like to offer for the good.

5. If a well-informed buyer chose the Name Your Own Price seller and his offer was not successful (because it was lower than the threshold price), the buyer can purchase the good from the posted-price seller, if he entered the market.

6. The four follow-up buyers learn which sellers are available. If only one seller is available, the follow-up buyers have the possibility to pay search costs of **10 points** to learn whether the respective other seller has entered the market or not. In case of the posted-price seller, the buyers also learn his posted price.

7. The follow-up buyers decide whether to buy the good and if yes, where to buy it. If a buyer chooses the Name Your Own Price seller, he also needs to decide which price he would like to offer for the good. If this offer was not successful, the buyer can decide whether he wants to buy the good from the posted-price seller, provided that the posted-price seller is available.

8. Buyers learn their income that they have earned in the current period.

   posted-price sellers learn how many buyers bought at their posted prices, how many buyers did not purchase, as well as their own income.

   Name Your Own Price sellers learn the price offers submitted by each of their buyers and whether these offers were successful. In addition, Name Your Own Price sellers are informed about the number of buyers who did not purchase, as well as their own income.

End of a Period

At the end of each period, sellers and buyers are separated and two sellers (one posted-price seller and one Name Your Own Price seller) are assigned to six new, randomly selected, buyers. Hence, a market always consists of one posted-price seller, one Name Your Own Price seller, two well-informed buyers and four follow-up buyers. After 20 periods the experiment ends.

CALCULATION OF INCOME AT THE END OF EACH PERIOD

If none of the sellers have entered the market in a given period, both sellers and the respective six buyers in the market receive **0 points**.

WELL-INFORMED BUYER

Transacting with the Posted-Price Seller

If a buyer purchases from the posted-price seller, his income equals his valuation V minus the posted price P_P.

$$\text{Income} = \text{Valuation } (V) - \text{Posted Price } (P_P)$$

Transacting with the Name Your Own Price Seller

If a buyer purchases from the Name Your Own Price seller, his income equals his valuation V minus the price paid P.

$$\text{Income} = \text{Valuation } (V) - \text{Price Paid } (P)$$

No Transaction

If the buyer does not trade in this period his income equals **0 points**.

FOLLOW-UP BUYER

The income of the follow-up buyers is calculated like the income of the well-informed buyers. If a follow-up buyer has invested search costs, these costs will also be deducted.

SELLER

The income of a seller equals his revenue minus his costs.

- His revenue are the prices he determined: these are the posted prices multiplied with the number of sold goods for the posted-price seller, and the sum of offered prices that are greater than or equal to his price threshold for the Name Your Own Price seller.

- His costs are calculated by multiplying the unit costs (C) with the number of goods sold.

If a seller or a buyer incurs a loss in any given period, they have to use their previous profits or their initial endowment, respectively, to compensate for this loss.

Do you have any questions up to this point?

C

APPENDIX TO CHAPTER 3

C.1 ROBUSTNESS

C.1.1 *Full Sample*

The analyses presented in the main text are based on a restricted sample. As described in more detail in Section 3.4.2.1, 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 C.1, coefficients point in the same direction as in Table 3.4, and for female participants both the coefficient on *Attractiveness Range* and the interaction with *More Attractive* are significant.

C.1.2 *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 C.2 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.

C.1.3 *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

	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 C.1.—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)	(2)	(3)
	All	Male	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 C.2.—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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 C.3 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 3.5.2: 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 coefficient 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 C.4. Results are qualitatively in line with Table

	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 C.3.—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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	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 C.4.—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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.4, 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.

C.1.4 *Consideration Set Composition*

This section complements the analysis in Section 3.5.4. In order to evaluate whether there are differential effects of consideration set composition on context dependence, Tables C.5 and C.6 split the sample into male and female evaluators.

C.2 GENDER DIFFERENCES IN UTILITY

The results presented in Section 3.5.2 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 C.1 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.

C.3 UNIQUENESS OF THE REFERENCE LEVEL

This section complements the analysis presented in Section 3.5.3 in the main part.

Figure C.2 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 C.2a 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 be-

	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

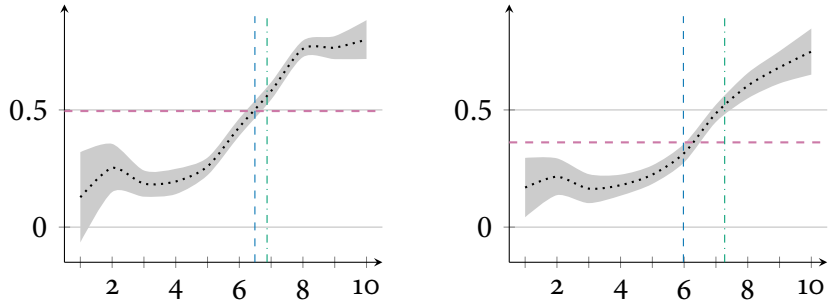
TABLE C.5.—Consideration Set Composition (Men)

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

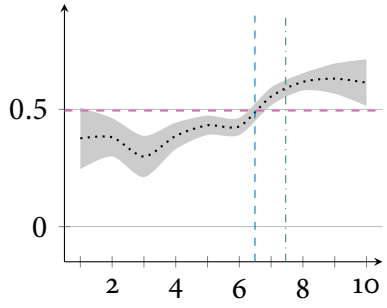
TABLE C.6.—Consideration Set Composition (Women)

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

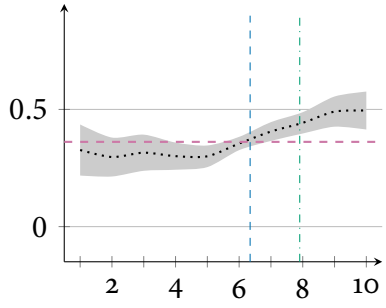


(A) Attractiveness (Men)

(B) Attractiveness (Women)



(C) Fun (Men)



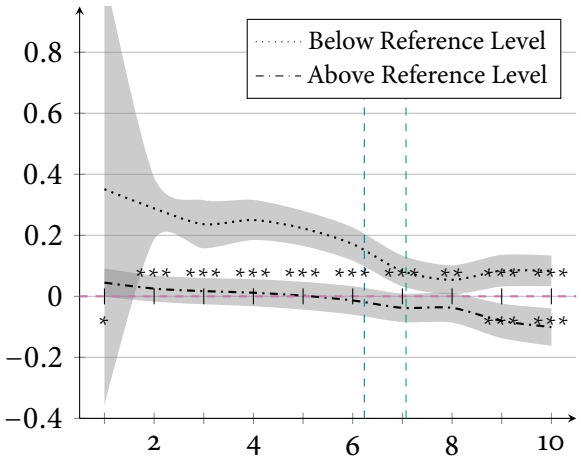
(D) Fun (Women)

FIGURE C.1.—Gender Differences in Utility

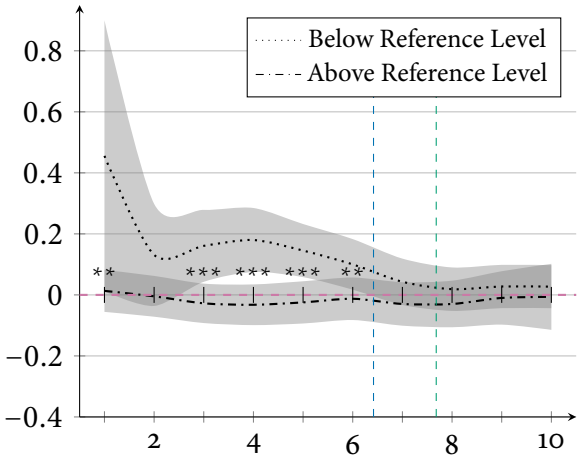
Notes: Vertical axis displays average predicted values of the dependent variable (decision to date) for different attribute values. Plotted values are derived from regressing the decision to date on all five attributes, including time and individual fixed effects. Marginal effects are obtained post estimation 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.

tween 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 3.4), 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 3.2: 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.



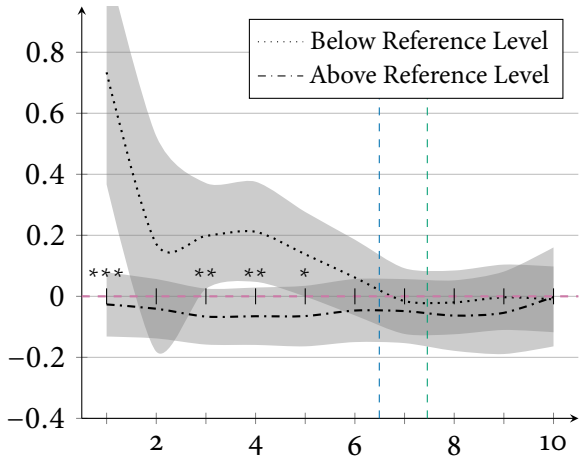
(A) Attractiveness



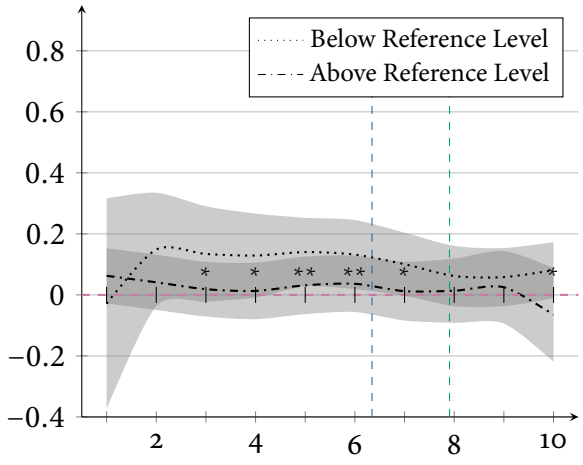
(B) Fun

FIGURE C.2.—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 C.3.—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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Part IV

BIBLIOGRAPHY

BIBLIOGRAPHY

Abbas, A. E. & Hann, I.-H. (2010). Measuring risk aversion in a Name-Your-Own-Price channel. *Decision Analysis*, *7*(1), 123–136.

Adams, C., Bell, A., Benton, T., Benston, S., Boggs, C., Davis, K., … Llorente, R., et al. (2011). *Critical theory and animal liberation*. Lanham: Rowman & Littlefield Publishers.

Amaldoss, W. & Jain, S. (2008). Joint bidding in the Name-Your-Own-Price channel: A strategic analysis. *Management Science*, *54*(10), 1685–1699.

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica*, *41*(6), 997–1016.

Anderson, C. K. (2009). Setting prices on priceline. *Interfaces*, *39*(4), 307–315.

Archer, E. J. (1962). Concept identification as a function of obviousness of relevant and irrelevant information. *Journal of Experimental Psychology*, *63*(6), 616.

Augoustinos, M., Walker, I., & Donaghue, N. (2014). *Social cognition: An integrated introduction*. London: Sage.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, *3*(3), 193–209.

Bartling, B., Weber, R. A., & Yao, L. (2015). Do markets erode social responsibility? *The Quarterly Journal of Economics*, *130*(1), 219–266. doi:10.1093/qje/qju031

Batra, R., Myers, J. G., & Aaker, D. A. (2009). *Advertising management*. London: Dorling Kindersley.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research*, *30*(5), 961–981.

Bem, S. L. (1981). *Bem sex-role inventory: Professional manual*. Palo Alto: Consulting Psychologists Press.

Bénabou, R. & Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics*, *126*(2), 805–855. doi:10.1093/qje/qjr002

Bhargava, S. & Fisman, R. (2014). Contrast effects in sequential decisions: Evidence from speed dating. *Review of Economics and Statistics*, *96*(3), 444–457.

Boone, L. E. & Kurtz, D. L. (2013). *Contemporary marketing*. Boston: Cengage Learning.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics*, *127*(3), 1243–1285.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, *121*(5), 803–843.

Bracha, A., Gneezy, U., & Loewenstein, G. (2015). Relative pay and labor supply. *Journal of Labor Economics*, *33*(2), 297–315.

Bray, J., Johns, N., & Kilburn, D. (2010). An exploratory study into the factors impeding ethical consumption. *Journal of Business Ethics*, *98*(4), 597–608. doi:10.1007/s10551-010-0640-9

Bushong, B., Rabin, M., & Schwartzstein, J. (2015). A model of relative thinking. *mimeo*.

Carrigan, M. & Attalla, A. (2001). The myth of the ethical consumer—do ethics matter in purchase behaviour? *Journal of Consumer Marketing*, *18*(7), 560–578.

Chernev, A. (2003). Reverse pricing and online price elicitation strategies in consumer choice. *Journal of Consumer Psychology*, *13*(1/2), 51–62.

Chesbrough, H. (2010). Business model innovation: Opportunities and barriers. *Long Range Planning*, *43*(2), 354–363.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: Theory and evidence. *The American Economic Review*, *99*(4), 1145–1177.

Christie, R., Geis, F. L., & Berger, D. (1970). *Studies in Machiavellianism*. New York City: Academic Press.

Chung, J. & Trivedi, V. U. (2003). The effect of friendly persuasion and gender on tax compliance behavior. *Journal of Business Ethics*, *47*(2), 133–145.

Costa, P. T. & McCrae, R. R. (1992). Neo PI-R professional manual. *Psychological Assessment Resources*, *396*, 653–65.

Cowe, R. & Williams, S. (2000). *Who are the ethical consumers?* Manchester: The Cooperative Bank.

Dawkins, R. (2009). *The god delusion*. New York City: Random House.

De Pelsmacker, P., Driesen, L., & Rayp, G. (2005). Do consumers care about ethics? Willingness to pay for fair-trade coffee. *Journal of Consumer Affairs*, *39*(2), 363–385.

Deckers, T., Armin, F., Kosse, F., & Szech, N. (2016). Homo moralis: Personal characteristics, institutions, and moral decision-making. *CESifo Working Paper*.

Del Barrio, V., Aluja, A., & García, L. F. (2004). Relationship between empathy and the big five personality traits in a sample of spanish adolescents. *Social Behavior And Personality*, *32*(7), 677–682.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, *47*(2), 315–372.

DePaulo, B. M. & Rosenthal, R. (1979). Telling lies. *Journal of Personality and Social Psychology*, *37*(10), 1713.

Dertwinkel-Kalt, M., Gerhardt, H., Riener, G., Schwerter, F., & Strang, L. (2016). Concentration bias in intertemporal choice. *mimeo*.

Descartes, R. (1996). *Meditations on first philosophy with selections from the objections and replies* (J. Cottingham, Ed. & Trans.). Cambridge: Cambridge University Press.

Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C., & Sirigu, A. (2009). Movement intention after parietal cortex stimulation in humans. *Science*, *324*(5928), 811–813.

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2005). Sources of openness/intellect: Cognitive and neuropsychological correlates of the fifth factor of personality. *Journal of Personality*, *73*(4), 825–858.

Ding, M., Eliashberg, J., Huber, J., & Saini, R. (2005). Emotional bidders: An analytical and experimental examination of consumers' behavior in a Priceline-like reverse auction. *Management Science*, *51*(3), 352–364.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *The American Economic Review*, *100*(3), 1238–1260.

Dolan, R. J. & Moon, Y. (2000). Pricing and market making on the internet. *Journal of Interactive Marketing*, *14*(2), 56–73.

Dufwenberg, M. & Gneezy, U. (2000). Price competition and market concentration: An experimental study. *International Journal of Industrial Organization*, *18*(1), 7–22.

Eisler, M. C., Lee, M. R., Tarlton, J. F., Martin, G. B., Beddington, J., Dungait, J. A., … Miller, H. (2014). Agriculture: Steps to sustainable livestock. *Nature*, *507*(7490), 32.

Falk, A. & Szech, N. (2013). Morals and markets. *Science*, *340*(6133), 707–711. doi:10.1126/science.1231566

Falk, A. & Szech, N. (2014). Diffusion of being pivotal and immoral outcomes. *mimeo*.

Falk, A. & Tirole, J. (2016). Narratives, imperatives and moral reasoning. *mimeo*.

Fay, S. (2004). Partial-repeat-bidding in the Name-Your-Own-Price channel. *Marketing Science*, *23*(3), 407–418.

Fay, S. (2009). Competitive reasons for the Name-Your-Own-Price channel. *Marketing Letters*, *20*(3), 277–293.

Fay, S. & Laran, J. (2009). Implications of expected changes in the seller's price in Name-Your-Own-Price auctions. *Management Science*, *55*(11), 1783–1796.

Fay, S. & Lee, S. H. (2015). The role of customer expectations in Name-Your-Own-Price markets. *Journal of Business Research*, *68*(3), 675–683.

Fehr, B., Samsom, D., & Paulhus, D. L. (2013). The construct of Machiavellianism: Twenty years later. *Advances in Personality Assessment*, *9*, 77.

Fehr, E. & Fischbacher, U. (2002). Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *The Economic Journal*, *112*(478), C1–C33.

Felthous, A. R. & Kellert, S. R. (1986). Violence against animals and people: Is aggression against living creatures generalized? *Journal of the American Academy of Psychiatry and the Law*, *14*(1), 55–69.

Filiz-Ozbay, E. & Ozbay, E. Y. (2007). Auctions with anticipated regret: Theory and experiment. *The American Economic Review*, *97*(4), 1407–1418.

Fischbacher, U. (2007). Z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.

Fisman, R., Iyengar, S. S., Kamenica, E., & Simonson, I. (2006). Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics*, *121*(2), 673–697.

Fisman, R., Iyengar, S. S., Kamenica, E., & Simonson, I. (2008). Racial preferences in dating. *The Review of Economic Studies*, *75*(1), 117–132.

Flynn, C. P. (1999). Animal abuse in childhood and later support for interpersonal violence in families. *Society and Animals*, *7*(2), 161–172.

Flynn, C. P. (2000). Why family professionals can no longer ignore violence toward animals. *Family Relations*, *49*(1), 87–95.

Ford, R. C. & Richardson, W. D. (1994). Ethical decision making: A review of the empirical literature. *Journal of Business Ethics*, *13*(3), 205–221.

Francione, G. (1995). *Animals, property & the law*. Philadelphia: Temple University Press.

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, *40*(2), 351–401.

Gabaix, X. & Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *The Quarterly Journal of Economics*, *121*(2), 505–540.

Gautier, P. A. & Klaauw, B. v. d. (2012). Selection in a field experiment with voluntary participation. *Journal of Applied Econometrics*, *27*(1), 63–84.

Geis, F. L. & Moon, T. H. (1981). Machiavellianism and deception. *Journal of Personality and Social Psychology*, *41*(4), 766–775.

Gennaioli, N. & Shleifer, A. (2010). What comes to mind. *The Quarterly Journal of Economics*, *125*(4), 1399–1433.

Gneezy, A., Gneezy, U., Nelson, L. D., & Brown, A. (2010). Shared social responsibility: A field experiment in Pay-What-You-Want pricing and charitable giving. *Science*, *329*(5989), 325–327.

Gneezy, A., Gneezy, U., Riener, G., & Nelson, L. D. (2012). Pay-What-You-Want, identity, and self-signaling in markets. *Proceedings of the National Academy of Sciences*, *109*(19), 7236–7240.

Gneezy, U., Imas, A., & Madarász, K. (2014). Conscience accounting: Emotion dynamics and social behavior. *Management Science*, *60*(11), 2645–2658.

Gordon, J. S. (1996). The chicken story: A century ago you'd eat steak and lobster when you couldn't afford chicken. *American Heritage*, *47*, 52–67.

Graziano, W. G., Habashi, M. M., Sheese, B. E., & Tobin, R. M. (2007). Agreeableness, empathy, and helping: A person × situation perspective. *Journal of Personality and Social Psychology*, *93*(4), 583.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer & V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen* (Vol. 63, pp. 79–93). Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.

Gruen, L. (2014). The moral status of animals. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2014).

Gullone, E. (2011). Conceptualising animal abuse with an antisocial behaviour framework. *Animals*, *1*(1), 144–160.

Hann, I.-H. & Terwiesch, C. (2003). Measuring the frictional costs of online transactions: The case of a Name-Your-Own-Price channel. *Management Science*, *49*(11), 1563–1579.

Harrison, G. W. & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *42*(4), 1009–1055.

Heidhues, P., Kőszegi, B., & Murooka, T. (forthcoming). Inferior products and profitable deception. *Review of Economic Studies*.

Heidhues, P., Kőszegi, B., & Murooka, T. (2016). Exploitative innovation. *American Economic Journal: Microeconomics*, *8*(1), 1–23.

Hinz, O., Hann, I.-H., & Spann, M. (2011). Price discrimination in e-commerce? An examination of dynamic pricing in Name-Your-Own-Price markets. *MIS Quarterly*, *35*(1), 81–98.

Hinz, O. & Spann, M. (2008). The impact of information diffusion on bidding behavior in secret reserve price auctions. *Information Systems Research*, *19*(3), 351–368.

Hitchens, C. (2008). *God is not great: How religion poisons everything*. New York City: Random House.

Holt, C. A. & Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review*, *92*(5), 1644–1655.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9*(1), 90–98.

Huber, J. & Puto, C. (1983). Market boundaries and product choice: Illustrating attraction and substitution effects. *Journal of Consumer Research*, *10*(1), 31–44.

Huffman, D. & Barenstein, M. (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*.

Imai, S. & Garner, W. R. (1965). Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology*, *69*(6), 596.

Jarvenpaa, S. L. (1990). Graphic displays in decision making—the visual salience effect. *Journal of Behavioral Decision Making*, *3*(4), 247–262.

Johnson, J. W. & Cui, A. P. (2013). To influence or not to influence: External reference price strategies in Pay-What-You-Want pricing. *Journal of Business Research*, *66*(2), 275–281.

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2012). The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience*, *7*(4), 393–402.

Kahneman, D. (1973). *Attention and effort*. Upper Saddle River: Prentice Hall.

Kalyanaram, G. & Winer, R. S. (1995). Empirical generalizations from reference price research. *Marketing Science*, *14*(3), 161–169.

Kant, I. (1981). *Lectures on ethics* (L. Infield, Trans.). Cambridge: Hackett Publishing Company.

Kelemen, L. E., Kushi, L. H., Jacobs, D. R., & Cerhan, J. R. (2005). Associations of dietary protein with disease and mortality in a prospective study of postmenopausal women. *American Journal of Epidemiology*, *161*(3), 239–249.

Kim, J.-Y., Natter, M., & Spann, M. (2009). Pay What You Want: A new participative pricing mechanism. *Journal of Marketing*, *73*(1), 44–58.

Kim, J.-Y., Natter, M., & Spann, M. (2010). Kish: Where customers pay as they wish. *Review of Marketing Science*, *8*(2), Article 3.

Kiousis, S. (2004). Explicating media salience: A factor analysis of New York Times issue coverage during the 2000 US presidential election. *Journal of Communication*, *54*(1), 71–87.

Kiousis, S. & Wu, X. (2008). International agenda-building and agenda-setting: Exploring the influence of public relations counsel on US news media and public perceptions of foreign nations. *International Communication Gazette*, *70*(1), 58–75.

Kőszegi, B. & Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, *121*(4), 1133–1165.

Kőszegi, B. & Rabin, M. (2007). Reference-dependent risk attitudes. *The American Economic Review*, *97*(4), 1047–1073.

Kőszegi, B. & Rabin, M. (2008). Revealed mistakes and revealed preferences. *The Foundations of Positive and Normative Economics: A Handbook*, 125–154.

Kőszegi, B. & Szeidl, A. (2013). A model of focusing in economic choice. *The Quarterly Journal of Economics*, *128*(1), 53–104.

Latané, B. & Nida, S. (1981). Ten years of research on group size and helping. *Psychological Bulletin*, *89*(2), 308.

León, F. J., Noguera, J. A., & Tena-Sánchez, J. (2012). How much would you like to pay? Trust, reciprocity and prosocial motivations in El Trato. *Social Science Information*, *51*(3), 389–417.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behavioral and Brain Sciences*, *8*, 529–566.

Locke, J. (1996). *Some thoughts concerning education: And, of the conduct of the understanding*. Cambridge: Hackett Publishing Company.

Loe, T. W., Ferrell, L., & Mansfield, P. (2000). A review of empirical studies assessing ethical decision making in business. *Journal of Business Ethics*, *25*(3), 185–204.

Loomes, G. & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, *92*(368), 805–824.

Lund, D. B. (2000). An empirical examination of marketing professionals' ethical behavior in differing situations. *Journal of Business Ethics*, *24*(4), 331–342.

Mak, V., Zwick, R., Rao, A. R., & Pattaratanakun, J. A. (2015). "Pay what you want" as threshold public good provision. *Organizational Behavior and Human Decision Processes*, *127*, 30–43.

Manzini, P. & Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, *82*(3), 1153–1176.

Marshall, B. & Dewe, P. (1997). An investigation of the components of moral intensity. *Journal of Business Ethics*, *16*(5), 521–529.

Masatlioglu, Y. & Nakajima, D. (2015). Completing incomplete revealed preference under limited attention. *Japanese Economic Review*, *66*(3), 285–299.

Masatlioglu, Y., Nakajima, D., & Ozbay, E. Y. (2012). Revealed attention. *The American Economic Review*, *102*(5), 2183–2205.

Matejka, F. & McKay, A. (2014). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *The American Economic Review*, *105*(1), 272–298.

McCarthy, R. F. & Bennett, R. E. (1986). Statutory protection for farm animals. *Pace Environmental Law Review*, *3*(2), 229–255.

McCrae, R. R. & Costa, P. T. (1997). *Conceptions and correlates of openness to experience*. London: Academic Press.

Muris, P., Meesters, C., & Timmermans, A. (2013). Some youths have a gloomy side: Correlates of the dark triad personality traits in non-clinical adolescents. *Child Psychiatry & Human Development*, *44*(5), 658–665.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making*, *6*(8), 771–781.

Nicholls, A. & Lee, N. (2006). Purchase decision-making in fair trade and the ethical purchase 'gap': Is there a fair trade twix? *Journal of Strategic Marketing*, *14*(4), 369–386.

Niederle, M. (2016). Gender. In J. H. Kagel & A. E. Roth (Eds.), *Handbook of Experimental Economics, Vol. 2*. Princeton: Princeton University Press.

O'Fallon, M. J. & Butterfield, K. D. (2005). A review of the empirical ethical decision-making literature: 1996–2003. *Journal of Business Ethics*, *59*(4), 375–413. doi:10.1007/s10551-005-2929-7

Olver, J. M. & Mooradian, T. A. (2003). Personality traits and personal values: A conceptual and empirical integration. *Personality and Individual Differences*, *35*(1), 109–125.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, *72*(6), 407–418.

Parducci, A. (1974). Contextual effects: A range-frequency analysis. *Handbook of Perception*, *2*, 127–141.

Pepitone, A. & DiNubile, M. (1976). Contrast effects in judgments of crime severity and the punishment of criminal violators. *Journal of Personality and Social Psychology*, *33*(4), 448–459.

Rabin, M. (2002). Inference by believers in the law of small numbers. *The Quarterly Journal of Economics*, *117*(3), 775–816.

Radtke, R. R. (2000). The effects of gender and setting on accountants' ethically sensitive decisions. *Journal of Business Ethics*, *24*(4), 299–312.

Regner, T. (2014). Social preferences? Google answers! *Games and Economic Behavior*, *85*, 188–209.

Regner, T. & Barria, J. A. (2009). Do consumers pay voluntarily? The case of online music. *Journal of Economic Behavior & Organization*, *71*(2), 395–406.

Regner, T. & Riener, G. (forthcoming). Privacy is precious: On the attempt of lifting anonymity on the internet to increase revenue. *Journal of Economics & Management Strategy*.

Ries, A. & Trout, J. (2001). *Positioning: The battle for your mind*. New York City: McGraw-Hill.

Ross, W. T. & Robertson, D. C. (2003). A typology of situational factors: Impact on salesperson decision-making about ethical issues. *Journal of Business Ethics*, *46*(3), 213–234.

Rothenhäusler, D., Schweizer, N., & Szech, N. (2015). Institutions, shared guilt, and moral transgression. *CESifo Working Paper*.

Sakalaki, M., Richardson, C., & Thépaut, Y. (2007). Machiavellianism and economic opportunism. *Journal of Applied Social Psychology*, *37*(6), 1181–1190.

Schmidt, K. M. (2009). The role of experiments for the development of economic theories. *Perspektiven der Wirtschaftspolitik*, *10*(si), 14–30.

Schmidt, K. M., Spann, M., & Zeithammer, R. (2015). Pay What You Want as a marketing strategy in monopolistic and competitive markets. *Management Science*, *61*(6), 1217–1236.

Shapiro, D. (2011). Profitability of the Name-Your-Own-Price channel in the case of risk-averse buyers. *Marketing Science*, *30*(2), 290–304.

Simonson, I. & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, *29*(3), 281–295.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690.

Sims, C. A. (2006). Rational inattention: Beyond the linear-quadratic case. *The American Economic Review*, *96*(2), 158–163.

Singer, P. (1995). *Animal liberation*. New York City: Random House.

Singhapakdi, A. & Vitell, S. J. (1990). Marketing ethics: Factors influencing perceptions of ethical problems and alternatives. *Journal of Macromarketing*, *10*(1), 4–18.

Singhapakdi, A., Vitell, S. J., Rao, C. P., & Kurtz, D. L. (1999). Ethics gap: Comparing marketers with consumers on important determinants of ethical decision-making. *Journal of Business Ethics*, *21*(4), 317–328.

Smith, V. L. (1976). Experimental economics: Induced value theory. *American Economic Review*, *66*(2), 274–279.

Smith, V. L. (2008). Experimental economics. In S. N. Durlauf & L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics*. Basingstoke: Palgrave Macmillan.

Sobel, J. (2010). Markets and other-regarding preferences. *mimeo*.

Soltani, A., De Martino, B., & Camerer, C. (2012). A range-normalization model of context-dependent choice: A new model and evidence. *PLoS Computational Biology*, *8*(7), e1002607.

Spann, M., Häubl, G., Skiera, B., & Bernhardt, M. (2012). Bid-elicitation interfaces and bidding behavior in retail interactive pricing. *Journal of Retailing*, *88*(1), 131–144.

Spann, M., Skiera, B., & Schäfers, B. (2004). Measuring individual frictional costs and willingness-to-pay via Name-Your-Own-Price mechanisms. *Journal of Interactive Marketing*, *18*(4), 22–36.

Spann, M. & Tellis, G. J. (2006). Does the internet promote better consumer decisions? The case of Name-Your-Own-Price auctions. *Journal of Marketing*, *70*(1), 65–78.

Spann, M., Zeithammer, R., & Häubl, G. (2010). Optimal reverse-pricing mechanisms. *Marketing Science*, *29*(6), 1058–1070.

Straughan, R. D. & Roberts, J. A. (1999). Environmental segmentation alternatives: A look at green consumer behavior in the new millennium. *Journal of Consumer Marketing*, *16*(6), 558–575.

Sullins, E. S. (1989). Perceptual salience as a function of nonverbal expressiveness. *Personality and Social Psychology Bulletin*, *15*(4), 584–595.

Sun, Y., Li, S., & Bonini, N. (2010). Attribute salience in graphical representations affects evaluation. *Judgment and Decision Making*, *5*(3), 151–158.

Sun, Y., Li, S., Bonini, N., & Su, Y. (2012). Graph-framing effects in decision making. *Journal of Behavioral Decision Making*, *25*(5), 491–501.

Terwiesch, C., Savin, S., & Hann, I.-H. (2005). Online haggling at a Name-Your-Own-Price retailer: Theory and application. *Management Science*, *51*(3), 339–351.

Thompson, G. D. (1998). Consumer demand for organic foods: What we know and what we need to know. *American Journal of Agricultural Economics*, *80*(5), 1113–1118.

Thompson, K. L. & Gullone, E. (2003). Promotion of empathy and prosocial behaviour in children through humane education. *Australian Psychologist*, *38*(3), 175–182.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, *26*(1), 24–36.

Todorović, D. (2010). Context effects in visual perception and their explanations. *Review of Psychology*, *17*(1), 17–32.

Trabasso, T. R. (1963). Stimulus emphasis and all-or-none learning in concept identification. *Journal of Experimental Psychology*, *65*(4), 398–406.

Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.

Van der Zee, K., Thijs, M., & Schakel, L. (2002). The relationship of emotional intelligence with academic intelligence and the big five. *European Journal of Personality*, *16*(2), 103–125.

Wang, T., Gal-Or, E., & Chatterjee, R. (2009). The Name-Your-Own-Price channel in the travel industry: An analytical exploration. *Management Science*, *55*(6), 968–979.

Weatherford, L. R. & Bodily, S. E. (1992). A taxonomy and research overview of perishable-asset revenue management: Yield management, overbooking, and pricing. *Operations Research*, *40*(5), 831–844.

Wilkowski, B. M., Robinson, M. D., & Meier, B. P. (2006). Agreeableness and the prolonged spatial processing of antisocial and prosocial information. *Journal of Research in Personality*, *40*(6), 1152–1168.

Wise, S. (2014). *Rattling the cage: Toward legal rights for animals*. Boston: Da Capo Press.