JULIUS-MAXIMILIANS-UNIVERSITÄT
WÜRZBURG

INAUGURAL-DISSERTATION ZUR
ERLANGUNG DER DOKTORWÜRDE DER
HUMANWISSENSCHAFTLICHEN FAKULTÄT

The Dual Nature of Utility

Categorical and Comparative Evaluations in Economic Decisions

Vorgelegt von

Michael Zürn

aus Rothenfels

Würzburg, März 2015

Erstgutachter: Professor Dr. Fritz Strack
Zweitgutachter: Professor Dr. Peter Bofinger

Tag des Kolloqiums: 24. Juli 2015

Contents

List of Figures

List of Tables

Acknowledgements:

First of all, I want thank my family and friends for supporting me in so many ways. I missed out on one or the other occasion to meet you, but you were always with me. Mom and Dad, know that it has always been you who helped me calm down and buck up when times were difficult, it has been you who gave me the strength to do all this! Caroline, you added warmth and joy to this journey and I am incredibly grateful you are by my side.

Fritz, without your openness to a new field of research, none of this would have been possible. Thank you for sparking my interest in social psychology and for being a wonderful supervisor during this whole enterprise. Moreover, I greatly thank Peter Bofinger for giving me the opportunity of working in another discipline which greatly enriched my understanding of economics.

I also thank all my colleagues for making me feel welcome even when facing the imminent danger of an economic invasion. Special thanks go to Anand, Anita and Thorsten. The discussions I had with you so often helped me see more clearly. Some obstacles were organizational in nature, but with Rita and Roswitha on my side, they were instantly swept away. To all my student assistants: you have my sincerest gratitude and you know that it have been the hours you spent in the lab that build the foundation of all this. Nonetheless, Dani, you were the best. Thank you!

Of course, there are so many more of you who deserve to be mentioned here. You know that my gratitude cannot be and will not be restricted to a few lines on a page. Nevertheless, very special thanks go to Simon for designing this awesome book cover and to Daniel for all the inspiration and simply a great time!

In Erinnerung an meine Oma Resi.

Chapter 1

Introduction

Triggered by the bankruptcy of Lehman Brothers in 2008, a global financial crisis emerged which ultimately cascaded into today's troubles of the Greek economy. Not only did the economic crisis wreak havoc on the daily lives of millions, but it also revealed a crisis of economics. The mere fact that the crisis was largely unpredicted by the scientific community only adds up to the more severe conundrum that, according to standard theories, it should never have happened in the first place. After all, textbook economics largely advertises the efficiency of markets which implies that "bubbles" (i.e. unjustified price levels), as well as their inevitable bursts, do not exist because prices always reflect all available information (Lo, 2008).

In general, attempts to substantiate the efficient market hypothesis start with assumptions about the behavior of the market participants. While any form of consensus is atypical for economists, this "methodological individualism" is broadly accepted in the scientific community (see Simon, 1959). That is, any macroeconomic phenomenon (economic growth, inflation, unemployment, etc.) has to be explained by the interactions

of economic agents. The efficiency hypothesis assumes that these agents are "rational" and therefore strive to maximize some mysterious quantity called "utility" (Sen, 2008).

However, the reality of crises and bubbles has cast serious doubt about actual market participants' utility maximization. Instead, decision making in financial markets was quickly attested a certain "Irrational Exuberance" (Shiller, 2015). In particular, it was hypothesized that a rational maximization of utility might be impeded by the workings of "animal spirits" (Akerlof & Shiller, 2010; Keynes, 1939/2006). As a result of this reasoning, bankers' "greed" was often invoked as the driving force behind the deals that led to the financial crisis. This seems to suggest that if "greed" (and presumably other emotions) had been part of economic theorizing, the financial crisis could have been predicted and even prevented.

As much as emotions must be acknowledged as determinants of behavior, their sudden prominence in attempts to account for the financial crisis appears highly speculative and ad hoc. To be sure, the general idea of combining psychological theorizing with economic models about human behavior promises to be a fruitful direction of research. Nevertheless, while a certain class of economic phenomena (see Loewenstein & Lerner, 2003) has been correctly ascribed to an impulsive system (cf. Strack & Deutsch, 2004), the idiosyncrasies of reflective evaluations have been largely overlooked as a basis of utility. In the present investigation, we will adopt a psychological perspective to take a closer look at the concept of utility.

Therefore, the next chapter gives a historical review of utility theory to ground this investigation firmly in economic theorizing and to summarize the problems

with the current state of affairs. In the third chapter, we introduce a Dual Utility Model accounting for different psychological processes underlying evaluative judgments. The fourth chapter contains six experiments which test central predictions of the proposed Dual Utility Model. The last chapter will discuss the findings with regard to both, the standard notion of utility and the proposed Dual Utility framework.

Chapter 2

Developments in Utility Theory

The following remarks are constrained to the milestones in the development of utility theory which paved the way for the current crisis of economic thinking. That is, we will begin with the early foundations on which the theory rests up to the present day and continue with the methodological refinements that have nonetheless been made in the subsequent century. These changes ultimately led to the first big wave of empirical testing in the mid-20th century revealing major inconsistencies with the theory's predictions. Also, these "anomalies" were a starting point for further adjustments in theory and experimental practice.

The selective approach of this review may be justified by excellent reviews on the topic that are already available (e.g. Bruni & Sugden, 2007; Edwards, 1954; Simon, 1959; Sugden, 1991; Stigler, 1950a, 1950b).

2.1 Gossen's Laws

The beginnings of utility theory may be found in Bernoulli (1738/1954). In particular, the Swiss mathematician realized that the value of an item is not adequately captured by only considering its monetary price. As a thought experiment, he described how persons *should* evaluate monetary gambles according to their respective levels of wealth. One of the gambles, called the "Petersburg Game", consisted of subsequent coin tosses which pay 2^{n-1} ducats, where n denotes the round in which the coin falls cross upward for the first time. Despite the game's infinite expected value, Bernoulli speculated that probably no one would pay much more than twenty ducats to play it. Apparently, the expected value did not correspond to the willingness-to-pay for an uncertain prospect. Consequently, Bernoulli formulated the first principle of modern decision theory: behavior is determined by the subjective evaluation of each action's objective consequences. More precisely, "the value of an item must not be based on its price, but rather on the utility it yields. The price of the item is dependent only on the thing itself and is equal for everyone; the utility, however, is dependent on the particular circumstances of the person making the estimate" (Bernoulli, 1738/1954, p. 24).

At the same time, simply introducing the concept of utility could not explain why potential gamblers are unwilling to pay infinite amounts to play the Petersburg Game. Therefore, Bernoulli (1738/1954) hypothesized that utility does not increase linearly with the objective consequences, in this case the possible payoffs of the game. Specifically, "any increase in wealth, no matter how insignificant, will always result in an increase in

utility which is inversely proportionate to the quantity of goods already possessed" (Bernoulli, 1738/1954, p. 25). Consequently, in contrast to the expected value in terms of money, the expected utility of the gamble is finite if this assumption holds. That is, finite monetary values for the opportunity to play the game can be ascribed to the *diminishing marginal utility* of money. While the first 100 Euros may increase utility drastically, the 101st 100 Euros probably may have only a small impact on total utility.

Similar assumptions about the relation between the quantity of a good and changes in utility had also been maintained in utilitarian philosophy (e.g. Bentham, 1789/1979; Mill, 1848). In general, this theoretical tradition is characterized by the hypothesis that nature had enthroned pleasure and pain as the "two sovereign masters" over human behavior (Bentham, 1789/1979). Thus, decisions are made with regard to their utility which in turn depends on the hedonic experience caused by the decisions' consequences. Accordingly, the intensity of pleasurable and painful experiences was assumed to diminish as the experience is repeated (see Rawls, Freeman, & Schulte, 2008). In addition, the utilitarian "concrete deductive method" (Mill, 1843, 1848), where a decision theory is derived from basic assumptions about human cognition, became the methodological foundation of economic theorizing. To be sure, economic theorizing had been closely intertwined with utilitarian philosophy (Rawls et al., 2008) but the concept of utility was not refined into a descriptive theory of human behavior until the "Marginalist Era" in economics (Niehans, 1990).

Even though Jevons (1871), Menger (1871) and Walras (1874) are generally considered the fathers of the "Marginal Revolution", it was Gossen (1854) who

first formulated two central principles of modern economic decision theory.[1] In detail, *Gossen's First Law* postulated diminishing marginal utility such that "the magnitude of a given pleasure decreases continuously if we continue to satisfy this pleasure without interruption until satiety is ultimately reached" (English translation taken from Niehans, 1990, p. 189). Hence, the marginal utility of a good is assumed to diminish as its quantity rises. However, consuming more units of a good does not increase utility any further once a certain satiation level is reached. That is, satiation implies that marginal utility ultimately equals zero or may even become negative.[2] At the same time, economic theory often assumes non-satiation for mathematical convenience because this guarantees that expenditures equal income (see Yang, 2001).

If the First Law holds, *Gossen's Second Law* follows from maximizing utility with scarce resources. In particular, Gossen formalized all property or endowment an economic agent controls and potentially trades as a "bundle of goods" which also is the entity carrying utility. The Second Law states that a utility maximizing agent will choose an allocation of the initial endowment where the marginal utilities of all goods in the bundle, standardized by each good's respective price,

[1]The "fathers" of the Marginal Revolution all arrived independently at Gossen's conclusion. See Stigler (1950b) or Schumpeter (1954/2009) for details about the individual works of each author.

[2]In fact, the sign of marginal utility (i.e. whether it is positive or negative) defines the very notion of a "good". That is, any product or service is a *good* if and only if its marginal utility is positive. Because marginal utility is subjective, one specific object may be a "good" for one person but an "ungood" (i.e. negative marginal utility) for another. For example, milk is probably a good for many people but definitely an ungood for those who are lactose intolerant (in a related vein see Ariely, Loewenstein, & Prelec, 2006).

are equal.[3] Importantly, marginal utilities and prices both refer to certain quantities of a good (e.g. 10 Euros for 1 Liter of wine).[4] Therefore, the standardization ensures that the equality dictated by the Second Law applies to quantities which are mutually exchangeable in the market. Hence, Gossen's Second Law prescribes a balanced composition of the consumption bundle such that "at the end" (i.e. when all income has been spent) each good contributes the same utility per Euro.

To be sure, consuming any good with a positive marginal utility increases total utility. Unfortunately, every time we make a decision to spend a scarce resource on something we necessarily decide against something else.[5] If we decide to see a movie, we cannot at the same time go to a concert. Also, even if consuming an additional glass of wine yields a positive marginal utility for the agent, he does not maximize utility if the alternative consumption of some sausage has a higher marginal utility. Hence, under scarcity, positive marginal utility is not sufficient to maximize total utility because a scarce resource spend to consume one good cannot be spend on another. The implications

[3]Formally, the Second Law consists of the first order conditions of a constrained utility maximization problem, while the First Law secures the second order condition for a maximum of utility. For a complete mathematical description of such a maximization problem see for example Yang (2001, pp. 71).

[4]In fact, every price is a ratio of quantities (wine for sausage, wine for a certain amount of currency) that can be exchanged for each other.

[5]The one resource which is ultimately always scarce is time. Therefore, Gossen (1854) chose the allocation of lifetime to different activities as an application of his theory. Also, the popular saying "There is no such thing like a free lunch" refers to the dilemma that even a free lunch costs time to eat. As a consequence, virtually every human decision involves scarcity even though decision makers probably are not always aware of it.

of this observation reach far. In general, economists consider transactions as means to alter the compositions of the bundles. That is, every transaction benefits the decision maker according to the marginal utility of the goods received but also involves costs equivalent to the marginal utility of the goods given up. Economic theory refers to the marginal utility of a rejected option as *opportunity costs* (Green, 1894; von Wieser, 1889, 1914/1924).[6] Therefore, "[c]ost is always a reflection of utility" (Niehans, 1990, p. 233) and the use of a resource is costly only in the sense that some other use must be foregone.[7] Concerning the decisions of economic agents, each transaction is evaluated by a comparison of marginal utility and opportunity costs.

Put differently, Gossen's Second Law describes the end state of a hypothetical arbitrage process where differences between marginal utilities and opportunity costs are eliminated by shifting resources from one good to another. Usually, realizing arbitrage is defined as taking advantage of price differences for the same good in different markets by buying low and selling high. From the present perspective, price differences refer to the inequality of marginal utility and opportunity costs. That is, buying low refers to a shift in resource allocation towards a good where the first outweigh the latter. By selling high, however, prices decrease due to the increased supply which diminishes

[6] More specifically, opportunity costs are defined as the marginal utility of the best alternative option which can be realized (see also Buchanan, 2008).

[7] Certainly, costs are commonly understood as the price which has to be paid for a good but as far as economic theory is concerned this is only part of the truth. In fact, these overt costs determine opportunity costs together with marginal utility of the alternatives in a multiplicative fashion. For example, if overt costs are high, then more alternatives and their respective marginal utilities have to be foregone.

marginal utilities until they ultimately equal opportunity costs. Consequently, any arbitrage opportunities are exceeded and the decision maker would be indifferent between the alternative courses of action.[8]

To a large degree, the concept of indifference first gained prominence in economic theorizing because Edgeworth (1877) generalized the notion of utility such that the utilities of two goods were no longer deemed independent of each other (for details see Creedy, 2008; Stigler, 1950b). For example, this generalization took into account that some goods are only useful in combination with others, e.g. like left and right shoes. Nevertheless, this development rendered previous methods of analyzing decision making unfeasible (Edwards, 1954; Stigler, 1950b). Therefore, Edgeworth (1881) introduced *indifference curves* to make predictions about human conduct. In particular, an indifference curve combines all bundles of two goods that yield the same level of total utility. Obviously, any decision maker

[8]For example, consider a vintner who possesses an ample endowment of wine but has no sausage to complete his meal. However, he might engage in a transaction with the butcher and barter wine for sausages according to the exchange conditions, i.e. the overt cost of the sausage in terms of wine. Moreover, if marginal utility diminishes, the initial composition of the vintner's bundle (a lot of wine but no sausage) might not be optimal. That is, while the marginal utility of wine is probably low due to the high quantity, the sausage's marginal utility might be quite high because the currently consumed quantity is very low. Therefore, the opportunity costs of sausage (i.e. the marginal utility of wine) might be lower than its marginal utility. Consequently, this would initiate a shift in the vintner's allocation giving up some of his wine in exchange for sausages. At the same time, the scope of this shift is constrained by the comparison of the sausage's marginal utility and opportunity costs because it will stop if the first outweighs the latter. In fact, if both are equal, the composition of the bundle is balanced according to Gossen's Second Law and the vintner will not engage in any further transactions.

would be indifferent between all bundles on the same indifference curve. At the same time, for all transactions along an indifference curve, marginal utility must always equal opportunity costs because total utility must remain unaffected by definition.[9] In a nutshell, the indifference curve constrains the transactions a utility maximizing agent is willing to accept. Moreover, Edgeworth's indifference curves were the basis for an important turn in the history of economic thought.

2.2 The Paretian Turn

For the pioneers of the Marginalist Era, there was little doubt about the the practice of deriving marginal utilities, indifference curves and in fact a whole decision theory from postulated utility functions. However, during the period between 1880 and 1900 the empirical foundations and interpretation of utility started to face growing suspicion of the scientific community. In the light of the "Methodenstreit", which generally revolved around the use of empirical data in economics (Schumpeter, 1954/2009), concerns were raised about the validity of the utilitarian assumptions underlying the concept of utility. In a related vein, there have been discussions about the representation of utility on a cardinal or ordinal scale. While Schumpeter (1954/2009) clarifies that utility was initially seen as a cardinally measurable quantity, the demise of the philosophical and psychological foundations of utility led economists towards an ordinal interpretation. This critical intellectual climate supported the second theoretical revo-

[9] Furthermore, Edgeworth (1881) formalized this statement by showing that the slope of the indifference curve is given by the ratio of marginal utility and opportunity cost (for details see e.g. Yang, 2001).

lution in economics which Bruni and Sugden (2007) have termed the "Paretian Turn".

Pareto (1909/1971) promoted "the dismal science" as a discipline independent of psychology (Bruni & Sugden, 2007). Therefore, he followed the tradition of Mill's deductive method, but required that the laws from which the decision theory had to be deduced should be based on a *proto-behaviorist* approach. In detail, utility should be inferred from the empirical observation of choices (Bruni & Sugden, 2007). More specifically, he suggested that utility functions may be constructed by identifying the "neutral" transactions which would leave the agent indifferent. Ultimately, Pareto (1909/1971) attempted to show that it is possible to reverse engineer utility functions from "indifference maps" which would emerge from combining all those neutral transactions. In sum, Pareto's reformulation of economic decision theory stated that mathematically integrating observable indifference maps yields an ordinal concept of utility. Alas, he was not able to formulate the adequate constraints about the consistency of the observed transactions which would guarantee the existence of such an ordering.

Despite Pareto's failure to formulate ordinal utility theory in every detail, he had already realized that "marginal utility [...] was unnecessary; the ratios of marginal utilities were all that mattered" (Niehans, 1990, p. 264).[10] Following this idea, Hicks and Allen (1934a) replaced marginal utility by the *marginal rate of substitution* (*MRS*). Specifically, the *MRS* is defined as the *ratio* of marginal utilities and is therefore preserved by any increasing transformation of the under-

[10]In fact, marginal utility is meaningful only if the units of utility are interpretable, that is, if utility is measurable on a cardinal scale.

lying utility function (see Yang, 2001).[11] At the same time, the *MRS* had to be derived from hypothetically observable demand functions. That is, stable relationships describing how the consumed quantity of a good varies with changes in prices and income. However, as had already been noted by Pareto, the ordinal preference field may not exist if demand functions do not satisfy certain criteria (see Katzner, 2008). In fact, this "integrability problem" occupied economic thinking for decades. As a consequence, a substantial amount of theorizing focused on the formulation of existence conditions for utility functions recovered from integrating a system of demand functions (Allen, 1932; Antonelli, 1886; Fisher, 1892/2007; Hicks & Allen, 1934a, 1934b; Slutsky, 1915/1953).

In direct response to the Hicks-Allen reformulation, Samuelson (1938) criticized the new theory for partly relying on intangible assumptions similar to the decision theory prior to the Paretian Turn. In detail, Samuelson admonished that "we cannot know the behaviour of ratios of marginal utilities" (Samuelson, 1938, p. 61). In fact, Hicks and Allen had introduced a constraint on the transactions such that "the more we substitute Y for X, the greater will be the marginal rate of substitution of Y for X" (Hicks & Allen, 1934a, p. 57). Thus, the higher a level of consumption for one good, the smaller the required compensation in form of another good. However, these authors could justify this assumption of an increasing *MRS* only because "it leads to the type of demand functions in the market which seem plausible" (Samuelson, 1938, p. 61).[12]

[11] Interestingly, Hicks and Allen intended not just a change in terminology but hoped to avoid the ostensible pitfall of merely "talking about one marginal utility by itself" (Hicks & Allen, 1934a, p. 53).

[12] In particular, Hicks and Allen (1934a) argued that an increas-

As a consequence, Samuelson (1938) suggested to ground utility on the composition of the bundles chosen under certain budget constraints. More specifically, assume that two bundles A and B have actually been chosen, each under a specific budget constraint. Then, bundle A can be assigned a higher utility than bundle B if the actual expenditure for bundle A is at least equal to the hypothetical expenditure for bundle B. This hypothetical expenditure is calculated by multiplying the quantities of bundle B with the prices which were available when bundle A was chosen.[13] As a consequence, if the expenditure for bundle A is higher than the hypothetical expenditure for bundle B then this bundle could have been purchased under these circumstances, but was not. Therefore, if this condition holds, the first bundle was "selected over" the second. Samuelson showed that if these relations between expenditures hold, then an ordinal utility function rationalizing the choices of the bundles exists. Moreover, the constraints implied in these rela-

ing *MRS* implies convex indifference curves and therefore secures the second order conditions (a negative second derivative) of constrained utility maximization. In contrast, if this assumption about the behavior of the *MRS* does not hold the internal equilibrium associated with any revealed choice would be unstable because any such choice would then denote a minimum of utility. Therefore, this "Hicks-Allen Law" is the mirror image of Gossen's First Law in the reversed line of reasoning where the theory starts with choice and develops conditions under which constrained utility maximization may have led to these choices (for details see Katzner, 2008).

[13] Mathematically, it is assumed that independent of the units in which prices and income are denoted, any agent will consistently show a unique demand structure $\boldsymbol{x}^i$ for n different goods under any given budget constraint. Thus, expenditure is given by the sum of the unique quantities multiplied with the respective prices $E^A = \sum \boldsymbol{p^A x^A}$. Also, it is possible to combine the prices of one observation with the quantities of another to calculate a hypothetical expenditure $E^{B'} = \sum \boldsymbol{p^B x^B}$.

tions have since been known as the "Weak Axiom of Revealed Preference" (WARP) (Samuelson, 1938). Despite the change in terminology from "utility" to "preference", the meaning of an evaluative ordering remained the same. However, Samuelson's approach took the observer's perspective from which choices *reveal preferences* interpreted as an advantage in terms of utility.

The ordinal utility approach of Hicks and Allen was not entirely reconciled with Revealed Preference Theory until Houthakker (1950) postulated the "Strong Axiom of Revealed Preference" (SARP). Houthakker's SARP extends WARP to include transitivity for choices over more than two bundles. That is, if bundle A "reveals itself" to be preferred over bundle B and the same is true for bundle B over C, then bundle A must also "reveal itself" to be preferred over bundle C.[14] Consequently, Samuelson (1950) summarized that the sometimes tedious discussion about the existence of ordinal utility functions has been ended by Houthakker's contribution. That is, any agent whose demand functions imply a non-integrable preference field, and therefore the absence of an ordinal utility function, will also violate the SARP (see also Samuelson, 1953). Hence, the SARP is a necessary and sufficient condition to test whether choice data is compatible with utility maximization. Subsequently, Afriat (1967) further refined Houthakker's analysis and extended it to finite sets of data which was indispensable for empirical testing. Based on Afriat's work, Varian (1982) proposed the "Generalized Axiom of Revealed Preference" (GARP) which is less restrictive than SARP because it is compatible with empirical data suggesting "locally flat" indifference curves. That is, if an actor would be indif-

[14] However, Sen (1973) argued that WARP already logically includes SARP.

ferent between bundles lying on the same subsection of the budget line, SARP would be violated but GARP would still hold (for a contemporary summary of Revealed Preference Theory see Varian, 2006).

In sum, if empirical budgetary data satisfies GARP, then the choices may be "rationalized" by a some utility function. At the same time, such behavior may be described *as-if* it follows from utility maximization (M. Friedman, 1953). In general, the empirical validity of Revealed Preference Theory seems to be given for human subjects (for a selection of empirical results concerning GARP see Varian, 2006). Even more, Kagel et al. (1975) have shown that rats and pigeons exhibit consistent demand behavior in the face of compensated price changes. Also, Chen, Lakshminarayanan, and Santos (2006) have shown that even the choices of capuchin monkeys in an artificial currency system satisfy GARP. Therefore, with the ordinal utility function now being firmly grounded in actual behavior, Edgeworth's indifference curve analysis was solidified as the basis of economic decision theory.

2.3 Expected Utility Theory

However, in the subsequent line of theorizing, even the concept of indifference curves was deemed irrelevant to derive utility. Specifically, von Neumann and Morgenstern (1944/1955) proposed that "under the conditions on which the indifference curve analysis is based very little extra effort is needed to reach a numerical [cardinal] utility" (p. 17, squared brackets added). This simplified concept of utility climaxes in an assumption "according to which *all the individual strives for is fully described by one numerical datum*" [emphasis added] (von Neumann & Morgenstern, 1944/1955, p. 33). A

simple example given by the pioneers of the theory may suffice to illustrate the underlying approach:

> "Assume that an individual prefers the consumption of a glass of tea to that of a cup of coffee, and the cup of coffee to a glass of milk. If we now want to know whether the last preference - i.e., difference in utilities - exceeds the former, it suffices to place him in a situation where he must decide this: Does he prefer a cup of coffee to a glass the content of which will be determined by a 50%-50% chance device as tea or milk" (von Neumann & Morgenstern, 1944/1955, p. 18).

That is, von Neumann and Morgenstern (1944/1955) realized that the probability of the second glass' content is a numerical measure for the preference between its possible contents. For example, if the certainty of an acceptable glass of coffee is preferred to a 50%-50% chance of either drinking the beloved cup of tea or the disliked glass of milk, then the utility difference between tea and coffee is smaller than between coffee and milk. Therefore, by applying probabilities to any kind of consequence (e.g. bundles of goods) it is possible to derive a numerical measure of utility from observable choices between the resulting *gambles*.[15]

[15]Later, Morgenstern (1979) expressed the high level of pragmatism he and von Neumann showed toward their cardinal interpretation of utility: "What von Neumann and I have done was simply to straighten out some issues that were not resolved at the time of our writing. We needed a number for the payoff matrices in game theory. We were also aware of the debate regarding ordinal and cardinal utilities. Instead of merely postulating the existence of a number for the purposes of game theory, which we easily could have done, we decided that we could obtain one by looking at the basic fact of uncertainty and it took very little time to formulate our axioms and give the necessary motivation." (p. 181).

Similar to Revealed Preference Theory, choices must conform to a set of mathematical axioms to guarantee the existence of the above mentioned numerical value (most prominently von Neumann and Morgenstern (1944/1955); Marschak (1950) and Savage (1954)).[16]

Most importantly, if the choices conform to the axioms, the value of all outcomes can be represented on a cardinal scale were a higher number reflects a greater utility (see also M. Friedman & Savage, 1952). Therefore, decisions under uncertainty can be modeled as maximizing the mathematical expectation of the utilities occurring in each possible event, i.e. the product of probabilities and values (e.g. M. Friedman & Savage, 1948). This Expected Utility Theory became the predominant model of choice under uncertainty in modern economics (Machina, 2008) and at the same time bears striking similarities to Bernoulli (1738/1954) who also studied gambles to derive a theory of utility.[17]

2.4 Anomalies

In the mid twentieth century, utility theory was confronted with an increasing number of *anomalies*. In fact, very shortly after the first sets of axioms had been

[16]To some degree, the axioms are shared with Revealed Preference Theory (completeness and transitivity), however additional requirements were introduced as well (e.g. the substitution axiom Marschak, 1950). A discussion of the differences and commonalities between the axiomatic systems is beyond the scope of this review but the interested reader may be referred to Malinvaud (1952), Samuelson (1952) and Sen (1973).

[17]Concerning the matter of uncertainty, the model has been axiomatized under conditions of both probabilistic and event-based uncertainty (for details see Machina, 2008). Furthermore, it is irrelevant whether uncertainty refers to probabilities regarding future "states of the world" (Savage, 1954) or whether it reflects strategic considerations from Game Theory.

published, evidence was presented that experimental subjects systematically violated them. Often, the axioms were characterized as "appealing" or "compelling, from a normative point of view" (Starmer, 2000, p. 334) which led to the conviction that no "rational" decision maker would willingly violate them if they were presented explicitly. As a consequence of this strong intuitive appeal, the axioms were subsequently treated not only as normative or prescriptive rules for "correct" behavior, but also as descriptive principles to predict decision making (see for example M. Friedman & Savage, 1952; Marschak, 1950; Sugden, 1991). Only in this latter meaning, they were able to constitute "anomalies" which are defined by the mismatch between a concrete prediction and empirical data. Importantly, the axioms never prescribe or predict a specific course of action for one instance of choice, but only require *consistency* over several choices.

One of the earliest demonstrations of inconsistent choices between monetary gambles was published by Allais (1953). In detail, Allais proposed several pairs of gambles which were designed such that a preference between two gambles implies a specific preference in a subsequent choice between two different gambles *if* the axioms of EUT hold. Most prominently, Kahneman and Tversky (1979) have performed experimental tests with gambles similar to those proposed by Allais (1953). For example, subjects first chose between two gambles, one with an 80% chance of winning \$4000 and the other with a certain win of \$3000. After their decisions, participants were asked to decide between two new gambles. This time, one gamble offered a 20% chance of winning \$4000, while the other offered a 25% chance of winning \$3000. An agent with rational or consistent preferences must select the first gamble in

the second round *if* he selected the first gamble in the first round because both pairs are identical except that the probabilities are quartered in the second round. However, participants exhibited a strong preference for the second gamble in the first round, but a preference for the first in the second round. That is, empirical preferences are non-linear in probabilities (Tversky & Kahneman, 1992) which violates the independence axiom (Marschak, 1950) and hence the assumption of a consistent preference ordering. At the same time, violations of consistency contradict the ideal of rational choices (Sen, 2008) and are therefore labeled "irrational".

In addition, the axiomatic systems imply certain assumptions concerning the methodology of testing the predictions of the theory. In particular, *procedure invariance* holds that utility is independent of the method of elicitation (Tversky, Sattath, & Slovic, 1988). For example, choice should reveal the same utility ordering as stating a willingness-to-pay for each option. That is, the option which is assigned the highest monetary value should also be chosen. Nonetheless, numerous studies found evidence for *preference reversals* between different methods to measure utility. For example, Lichtenstein and Slovic (1971) showed that a lottery with a high probability of a small gain (P-bet) was preferred in choice, but a second lottery with the same expected value but a low probability of a larger gain ($-bet) was assigned higher monetary values. Therefore, different elicitation procedures give rise to systematically different evaluations of the options. In a related vein, axiomatic utility theories presuppose *description invariance*. That is, the utility of an option should be independent of the way it is presented or "framed". Specifically, the true content of a glass does not change

if one person tells you that half of it had been spilled and another person ensures that half of it remained contained. Nevertheless, the first framing may elicit very different evaluations than the second. More drastically, Tversky and Kahneman (1981) showed strong framing effects using a short story about an "unusual Asian disease" which could be contained by two different programs varying in the level of risk. The twist of the study consisted in either describing both programs in terms of lives saved, which led most participants to prefer the less risky option, or in terms of lives lost, which revealed a general preference for the more risky option. Therefore, even if the options were identical, subjects' evaluation was sensitive to the ways they were described.

The list of anomalies that have been documented over the last decades is long and will not be reviewed at this point (for an overview see Thaler, 1994). Instead, our focus will shift towards the most prominent attempts to explain and remove the deficiencies of utility theory.

2.5 Remedies

The emergence of systematic violations of traditional utility theory has stimulated a vast literature attempting to explain these anomalies. Simultaneously, utility theory itself has been modified in the light of these explanations in order to correct for its deficiencies. To some degree, psychological processes have been recognized in these "remedies". Interestingly, Pareto (1909/1971) banned psychology from the territory of economics but also anticipated the lines of reasoning from which most support for the utility paradigm has emerged. In particular, while describing his "science

of logical action", which is based solely on observable behavior, he realized that further constraints on the basic empirical facts were necessary to ensure the validity of the data. Firstly, his "instrumentality criterion" excluded all behavior which is not instrumental in the sense that it is guided by a stable preference ordering. Therefore, impulsive actions were not considered a valid basis to infer utility.[18] Secondly, Pareto realized that only if behaviors are repeated many times, the feedback from trial-and-error learning allows a sequential adjustment of the decisions. Only after choices have stabilized, they are a viable basis to reveal the underlying utility ordering. The following subsections will shortly describe these two answers to the challenges of utility theory.

Impulsivity

To a certain degree, labeling theories featuring some form of utility maximization as *Rational Choice Theories* may have been suggestive for citing impulsive or emotional forces as the cause of "irrational" behavior. Certainly, this semantic construction gave rise to fruitful insights into a specific type of anomalous behavior which will be summarized under the term "impulse anomalies". In general, this class of anomalies is characterized by assuming that the decision maker has a consistent utility ordering available, but fails to translate this valuations into action (Loewenstein, 2000). For example, most people would probably assign a high utility to maintaining a certain standard of living dur-

[18] In addition, behavior guided by social norms is excluded by instrumentality criterion. Yet, addressing the extensive psychological research on this topic is beyond the scope of this enterprise. However, the experiment described in Section 4.1 to some degree addresses this issue.

ing their retirement. However, many of them may at the same time fail to save money to secure the availability of the necessary funds later on. That is, not saving for the cherished retirement reveals an inconsistency between behavior and utility. In general, time inconsistencies form an important subtype of these anomalies. A phenomenon which is often cited as an example of such inconsistencies is *hyperbolic discounting*. In detail, Rational Choice Theory often assumes that decision makers are impatient an prefer immediate to delayed consumption. In order to incorporate this feature of decision making, intertemporal utility functions include a constant discount factor to adjust the present value of future consumption. In line with this formal description, empirical investigations of intertemporal decisions show that some people are nonetheless impatient and prefer one apple today to two apples tomorrow. In contrast, almost everyone is patient in the future and prefers two apples in one year and a day to one apple in one year (Thaler, 1981). However, traditional discounting approaches hold that the reduction of the present value depends only on the delay itself which is the same in both decisions. As a consequence, such findings suggest that the discount factor increases as payday approaches. That is, temptation is harder to resist if it is right under one's nose.

In a related vein, a large literature on "impulse purchases" has been established (e.g. Baumeister, 2002; Loewenstein, 1996; Loewenstein & Lerner, 2003). On a basic level, researchers agree that some stimuli trigger an urge to behave in a certain way even though it is at odds with some personal guideline. Furthermore, immediately experienced affective qualities have been ascribed to these behavioral tendencies (Loewenstein, 1996). Simultaneously, it is obviously possible to re-

sist temptation if certain criteria are met. For example, Hofmann, Rauch, and Gawronski (2007) showed that participants who had set some dietary restraints for themselves consumed less candy than subjects who did not hold any dietary standards. In contrast, if cognitive control resources were depleted, personal standards did not predict consumption (for details see Hofmann, Friese, & Strack, 2009). Therefore, resisting temptation requires the availability of certain cognitive resources.

Not surprisingly, the explanations for these phenomena have been drawn along the lines of self-control. In particular, a duality mirroring recent dual process models from social psychology (e.g. Alos-Ferrer & Strack, 2014; Strack & Deutsch, 2004) has been incorporated to account for the affective and impulsive forces leading agents off the path of reason. Specifically, these models consist of a duality of qualitatively different cognitive processes which transform perceptual inputs into behavioral responses. First, via an impulsive route, the activation of associated mental concepts spreads through an associative network eventually activating behavioral schemata. Second, via a reflective route, perceptual inputs are categorized, evaluated and ultimately translated into a goal which guides behavior (for details see Strack & Deutsch, 2004). The resulting *Dual Self models* generally formalize self-control issues where "irrational" decisions are attributed to a lack willpower (see also Simon, 1955, 1983). For example, Fudenberg and Levine (2006) have proposed a dual-self-model where a short-run self, guided by short-run utility, plays a strategic game against a long-run self exerting control to maximize the present value of all short-run utilities. As a consequence, seemingly "irrational" behavior such as

hyperbolic discounting can be explained by these models.

To be sure, this field of research gathered substantial attention in the scientific community. For once, the prominence of studying how the "Siren's Call" affects human behavior was probably nurtured by the fascination for the new research methods employed in neuroscience (e.g. Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). In fact, "neuroeconomics" is often seen as the future of the discipline (e.g. Glimcher & Fehr, 2013). In addition, insights from research showing that decisions are often made in discord with the subject's best interest constitute the basis for inexpensive political interventions. While past programs often focused on educating the consumers to make more informed decisions, policy makers increasingly put their trust into "nudging" people towards better decisions (Thaler & Sunstein, 2008). Nevertheless, the role of information and learning for decision making has also been considered as a theoretical mean to recover the validity of utility theory.

Learning

Even in the light of the evidence against the idea of consistently maximizing utility, many experimental economists refused to give up the utility paradigm. As consequence, Grether and Plott (1979) have tried to clean up the original experimental setting of Lichtenstein and Slovic (1971) which showed systematic preference reversals. After all, the baffled economists trembled that "[t]he inconsistency is deeper than the mere lack of transitivity [...] It suggests that no optimization principles of any sort lie behind even

the simplest of human choices" (Grether & Plott, 1979, p. 623). Therefore, the authors thoroughly eliminated factors that would allow an explanation of the anomaly along the lines of economic theorizing (missing incentives, decreased risk aversion due to higher expected income, etc.) but failed to reduce the rate of preference reversals. A criticism which prevailed is the unfamiliarity with the goods involved in the experiments (see Braga & Starmer, 2005). Therefore, some economists argue in favor of the "Discovered Preference Hypothesis" (Binmore, 1999; Plott, 1996; Smith, 1994). Essentially, it states that "rational-choice theory is descriptive of the behaviour of economic agents who, through experience and deliberation, have learned to act in accordance with their underlying preferences; deviations from that theory are interpreted as short-lived errors" (Bruni & Sugden, 2007, p. 148). In line with Pareto (1909/1971), this approach assumes that after purchasing certain quantities of goods, the subsequent consumption may lead agents to either decrease or increase their demand for each good in the next period. This feedback loop is repeated until decision makers see no further need to change the allocation, i.e. when they see no further room for improvements. Hence, the discovered preference approach acknowledges that value is created by experience and deliberation but adopts a long-run perspective largely neglecting the psychological process of evaluation.

Nonetheless, Braga and Starmer (2005) concluded that the evidence for the Discovered Preference Hypothesis is rather mixed. In fact, some anomalies appear to decay with experience but others do not (see Cox & Grether, 1996; Kahneman, Knetsch, & Thaler, 1990; Loomes, Starmer, & Sugden, 2010; Shogren, Shin, Hayes, & Kliebenstein, 1994). In addition, the

necessary conditions for actually discovering preferences are repetition and adequate learning opportunities which seriously restricts the scope of this approach. As Bruni and Sugden (2007) have argued, economic science would have to choose whether to withdraw from conquered territories or to adjust its methodological foundation. In general, modern theories claim that the economic approach applies to a wide range of human behavior (e.g. G. S. Becker, 1965), even though many real life situations hardly justify the assumptions of Pareto's successors about appropriate opportunities to discover one's preferences.[19]

Generally, the psychological processes underlying utility have not received adequate attention in economic theorizing. As this section has shown, fragments of psychological mechanisms have been implicitly assumed but a diligent analysis has not been carried out. Therefore, we propose a deviation from the unitary concept of numerical utility and suggest distinguishing between two types of utility that are driven by different psychological processes. We propose that the present distinction affords a better understanding of the dynamics of financial transactions as well as seemingly irrelevant microeconomic decisions like the purchase of a discounted toaster.

[19]For example, Becker proposed that even decisions with extremely limited learning opportunities like marriage (G. S. Becker, 1973, 1974) may be described by his approach. Even more drastically, G. S. Becker and Murphy (1988) hypothesized that even drug addiction can be considered rational despite the fact that learning is systematically impeded and largely ineffective as the gathered experiential information is considerably biased.

Chapter 3

Specification of the Dual Utility Model

While the previous chapter sketched the corners of traditional utility theory, the following passages attempt to draw the picture of a more psychologically solid approach to the notion of utility. The starting point of this enterprise is the conceptualization of utility as the result of an evaluative judgment. To some degree, economists from the Austrian School already regarded the subjective nature of value as the result of evaluative judgments generated by the economic actors (Menger, 1871). Unfortunately, the prevailing tradition of economic thought obscured the cognitive processes leading to these evaluations. Most drastically, the simplified concept employed in Expected Utility Theory reveals the interpretation of value as a directly perceivable attribute of a good. For example, von Neumann and Morgenstern (1944/1955) already compared their own method of measuring utility with physicists' definition of an absolute zero in the measurement of temperature (see also M. Friedman & Savage, 1952). As a consequence, the dominant conviction emerged that

"[i]t simplifies matters, and involves no loss in generality, to regard the alternatives open to the consumer unit as capable of being expressed entirely in terms of money or money income. [...] This permits us to consider total utility a function of money income alone." (M. Friedman & Savage, 1948, p. 288). Apart from recognizing a diminishing utility of money, this simplification establishes a theoretical equivalence of utility and money. Thus, monetary payoffs are generally deemed feasible stimuli to investigate any aspect of decision making. However, in contrast to the obvious possibility of "directly perceiving", i.e. reading the value of a monetary payoff, the evaluation of goods in general requires a judgment.

As a reaction to the numerous anomalies casting doubt on the existence of stable utility orderings, a line of research emerged which started with the assumption "that people's preferences are often constructed in the process of elicitation" (Foreword of Lichtenstein & Slovic, 2006). In a related vein, Tversky and Thaler (1990) proposed three different views on the nature of value. First, decision makers might directly perceive values like temperature. Obviously, this view captures the approach adopted by researchers in the tradition of von Neumann and Morgenstern (1944/1955). Second, decision makers might store the evaluation in memory and recall them in the instance of choice. Largely, this view corresponds to the discovered preference approach (Plott, 1996) where evaluations are known from prior experience. Finally, the third view on values interprets them as the result of a psychological construction. Accordingly, "preference construction" (Lichtenstein & Slovic, 2006; Slovic, 1995) matches Menger (1871) in so far as utility is seen as the result of an evaluative judgment. In general, psychologists agree

that any judgment is always made in a specific context. That is, the outcome of a judgment reflects what comes to mind when it is formed. For example, Strack, Schwarz, and Gschneidinger (1985) showed that evaluations of life satisfaction were affected by the mental content which was made more available (see also Schwarz, Bless, et al., 1991; Schwarz, Strack, & Mai, 1991; Strack & Martin, 1987). Unfortunately, despite the considerable body of social psychological research about the dynamics of judgment, few attempts have been made to explain how preferences are constructed (for some exceptions see Simonson, 1989; Simonson & Tversky, 1992; Slovic, Griffin, & Tversky, 1990; Tversky & Simonson, 1993).

In reviewing an extensive body of research, Simonson (2008) concluded that the evidence in favor of preference construction often suggests that subjects gravitate towards comparative evaluations if the decision at hand allows it. In addition, however, certain inherent preference components which "typically relate to categorical, non-quantitative aspects (e.g., the taste of beef jerky, a motion sensitive videogame remote)" (Simonson, 2008, p. 20) also contribute to the construction of evaluative judgments. Interestingly, during the developments of utility theory categorical and comparative evaluations were both deemed important for economic decision making. However, both types of judgment have never been treated as distinct psychological processes. Therefore, the next section will introduce a Dual Utility Model elaborating on categorical and comparative features in the existing concept of utility by integrating it with research from social psychology.

3.1 The Utilitarian Duality

"Everything is relative." This essentially Einsteinian statement is often abused as a safety net for some line of argument. Also, it has often been argued that any form of information processing ubiquitously involves some kind of comparison (e.g. Mussweiler & Epstude, 2009). Even more proponents may be recruited for the claim that evaluative processes particularly rely on comparisons (e.g. Ariely & Jones, 2008; Hsee & Zhang, 2010). On a phenomenological level however, non-comparative evaluations appear to occur quite often. For example, the aesthetic value of a Picasso can be judged without comparing it to a Manet. In general, comparing pleasures and pains - with each other and among each other - is most certainly possible and pervasive, but it is not necessary to generate an evaluation of the underlying experience.[1]

Furthermore, conceptualizing a vast array of cognitive operations as comparisons runs the risk of diluting the concept itself. For instance, one might argue that comparisons and categorizations establish a continuum of cognitive operations. That is, a categorization could be framed as a rudimentary comparison of an object with the specifications of a category constituting a prototype (e.g. Bruner, 1957; Rosch & Lloyd, 1978). Like all judgments, the act of categorization is a statement about the relation between two cognitions, i.e. it applies the relational schema of category membership (Strack & Deutsch, 2004). Also, every comparative statement applies a relational schema,

[1] At the same time, if evaluations are articulated in communication, comparisons posit a feasible mean to ensure a common ground and establish a frame of reference. That is, in order to communicate one's personal evaluation of a new SciFi movie it may be very helpful to compare it with a timeless classic like *Star Wars*.

but not all relational schemas are comparative. That is, there are qualitative differences between the relations established by categorization and those of comparison. In particular, categorizations are holistic statements about objects without reference to a specific dimension, whereas comparisons focus on specific attributes establishing a dimension of comparison. Therefore, if an object is categorized as a chair, then this applies to the whole entity. In contrast, comparing a particular object with the mental prototype of a chair may yield factual statements about a higher number of legs or the absence of a backrest. Because the following model is based on the psychological processes of evaluative judgments, the qualitative difference between these operations is reflected in a categorical and a comparative utility component. Specifically, we assume that different types of informational and motivational antecedents give rise to two types of utility.

Categorical Utility

The first type of utility results from a simple categorization and will be referred to as "categorical utility" (uCat). In general, the categorization of an object depends on the "match between the characteristics of the input and the specifications of the category" (Bruner, 1957, p. 148).[2] On a fundamental cognitive level, assigning a superordinate category to distinct objects or events allows to treat them in the same or a similar way (Rosch & Lloyd, 1978; Mervis & Rosch, 1981). Therefore, knowledge can be applied in situations even though they might not yet have been ex-

[2]Research about the psychological nature of categorizations suggests that categorical judgments are not always binary but may involve different degrees of representativeness that reflect the similarity of an object to a more abstract prototype (Rosch, 1973).

perienced (Bruner, 1957). Because this also applies to abstract evaluations which have been established in the past, they can be transferred to concrete objects upon encounter (cf. Strack & Deutsch, 2004). More specifically, these evaluations may be attitudes (Fazio, 1986) or they may stem from setting a goal (cf. Ajzen, 1991). For example, perceptually accessible attributes like the rectangular shape of a colored surface may elicit the categorization of an newly encountered object as a painting. Moreover, this categorization may activate knowledge about a decorative potential of the object. As a consequence, the painting would be evaluated positively if the attribute "decorative" has been pre-evaluated positively.

To be sure, some attributes like the taste of food or the sound of music may not rely on the activation of prior evaluations to generate a judgment about the value of the corresponding object. Instead, they may elicit an immediate feeling which can then serve as basis for evaluation (Schwarz & Clore, 1983). This line of research has recently been summarized as "Feelings-As-Information-Theory" by Schwarz (2011) who concluded that affective experiences often are a source of information in a great multitude of judgments. For example, looking at the composition of Picasso's take on *The Luncheon on the Grass* may elicit positive feelings which translate into a positive evaluation of the painting. Corresponding to this perspective, Edgeworth (1881) based his treatment of marginal utility on state of the art psychophysical research (Fechner, 1860/1964; Weber, 1846/2006; Wundt, 1874). A central conclusion from this early works in experimental psychology mirrors the assumption of diminishing marginal utility. Specifically, the intensity of a sensation elicited by a constant change

of the stimulus material diminishes as the total magnitude of the stimulus increases. Furthermore, if these diminishing sensations serve as affective information to generate categorical evaluations, the view of feelings as information blends naturally with this economic definition of marginal utility (cf. Bentham, 1789/1979; Edgeworth, 1881; Pareto, 1909/1971).[3]

In a related vein, the Discovered Preference Hypothesis (Plott, 1996) tentatively suggests that not further specified knowledge is generated from prior experiences and is later on applied to a good of the same category. In the tradition of economic thought, it may be assumed that the generated knowledge is hedonic in nature. At the same time, Kahneman, Fredrickson, Schreiber, and Redelmeier (1993) argue that hedonic experiences are often remembered in a biased fashion. In particular, the duration of an experience is largely neglected which elicits a better evaluation of longer unpleasant experiences if their intensity decreases towards the end. More specifically, Redelmeier, Katz, and Kahneman (2003) found evidence for this "peak-end rule" in colonoscopy patients who remembered treatments more positively if they extended the painful procedure by a period of milder pain. As a consequence, evaluations based on distorted memories reveal a serious problem for the Discovered Preference Hypothesis and also lead to categorical evaluations which might be disappointed later on.

Importantly however, judgments concerning uCat can also include semantic knowledge acquired during previous experiences. That is, the attributes of an ob-

[3]Interestingly, even the developers of the numerical concept of utility assert "that every measurement [of utility] - or rather every claim of measurability - must ultimately be based on some immediate sensation" (von Neumann & Morgenstern, 1944/1955, p. 16).

ject may be part of knowledge structures from which evaluations can be transferred. In detail, a central attribute of any human made artifact is its purpose. For instance, the purpose of a ladder is reaching elevated areas, maybe in order to hang up a decorative painting. Although the inference of a categorized object's purpose may already suffice to generate an evaluation, sometimes the purpose of an object may be part of the narrative in mental simulations (e.g. Adaval & Wyer, 1998; S. T. Fiske, 1993; Kahneman & Miller, 1986). Supporting this idea, Elder and Krishna (2012) investigated the interplay between different modalities in these simulations. In detail, the authors showed that participants who were presented with pictures of products facilitating the mental simulation of consumption (e.g. a bowl of soup with a spoon) evaluated the goods more positively than participants who saw pictures inhibiting the mental simulation of the consumptive act (e.g. the same bowl of soup without a spoon). To a certain degree, these simulations also create an affective scenery which may yield further information used in evaluative judgments (Kahneman, Ritov, & Schkade, 1999). Importantly, this conceptualization of uCat extends the discovered preference approach such that in addition to discovery by experience, which restricts the applicability of the hypothesis to repeated choices, discovery by simulation also constitutes a way to gather information about yet unfamiliar experiences. Unfortunately, the simulations or forecasts are often subject to biases. For once, predictions are often not accurate when decision and anticipated experience exhibit a "hot-cold empathy gap" (Loewenstein & Schkade, 1999; Van Boven, Dunning, & Loewenstein, 2000). In such situations, the current visceral state of the decision maker does not match the state during the subsequent

experience (see also Gilbert, Gill, & Wilson, 2002; Kahneman & Snell, 1992; Loewenstein, O'Donoghue, & Rabin, 2003). For example, a hungry shopper may purchase too much food. Also, Loewenstein, Nagin, and Paternoster (1997) showed that participants who were sexually aroused predicted more aggressive mating behavior (see also Ariely & Loewenstein, 2006). As a consequence, the uCat that is ascribed to a good depends on both the attributes of that good and the state of the actor, such as a need or a goal (see also Helson, 1964).

In sum, we propose that marginal utility can be conceptualized as uCat because both refer to the evaluation of a concrete state of the world. More specifically, marginal utility depends on the categorization of the additional good and the evaluation of its attributes.[4] To a certain degree, the very notion of marginal utility suggests that some comparative process is involved. Specifically, the concept refers to a *change* in utility caused by a *change* in consumption. However, from a psychological vantage point it is doubtful whether the evaluation necessarily draws on the current level of consumption as a standard of comparison. Nonetheless, comparisons are often the basis of evaluative judgments and must be reflected in specific utility component.

Comparative Utility

In contrast to the categorization of objects and the subsequent transfer of evaluative information, the second type of utility results directly from a comparison. Hence, "comparative utility" (uCom) depends on the

[4]Furthermore, the combination of these characteristics with the united attributes of the remaining bundle determines the type of the good as a substitute or a complement for the other goods in the bundle (cf. Hicks & Allen, 1934a; Slutsky, 1915/1953).

distance between a target and a standard. Standards may be deliberately recruited by the decision maker or provided by the context of the decision. Often, the selection of a standard determines the dimension in which the comparison occurs. That is, comparisons can be performed on a multitude of dimensions. Additionally, the interpretation of distances between target and standard hinges on the level of measurement available in a particular dimension. For example, comparing the pleasure of indulging the look at a Picasso with that elicited by a Manet is most likely an ordinal comparison. In contrast, comparing the size or the price of the paintings is a cardinal comparison.

Even though comparative processes are also relevant for factual judgments ("Picasso lived longer than Manet"), comparisons with an evaluative content are the basis of uCom. As a result, uCom may be conceptualized in terms of advantages and disadvantages inferred from positive and negative distances between a target value and a standard. For example, a CEO may evaluate an incoming job offer by comparing the salary of ten million Euro with the nine million Euro she earns in her current job. If the first offer is assigned a positive uCom due to a comparison in the monetary dimension, she will most likely switch jobs. As a first central implication of evaluative comparisons with different standards, uCom is not necessarily subject to satiation. That is, while diminishing marginal utility may be a feasible assumption concerning uCat, comparative judgments about a given object may change if another standard is selected. As a result, a person earning 60.000 may not judge the attractivity of the ten million job much higher than the attractivity of the other which could be interpreted as satiation. However, from the CEO's perspective, the advantage of the bet-

ter paid position may be obvious when comparing ten million to nine.

In general, microeconomic analysis focuses on transactions involving costs and benefits. Therefore, Thaler (1985/2008) proposed that transactions are psychologically represented as "mental accounts" (see also Shafir & Thaler, 2006; Thaler, 1980, 1999). In particular, each account consists of a balance sheet where the benefits oppose the costs. For example, the decision to take a certain job concerns the transaction of the employer's money in exchange for the employee's time. Taking the employee's perspective, time represents the cost on the balance sheet while the wage represents the benefit. That is, "[t]he accounting system provides the inputs to do both ex ante and ex post cost-benefit analyses" (Thaler, 1999, p. 184). However, mental accounting is very flexible concerning the dimension in which costs and benefits are compared. For instance, the above example of the CEO suggests that the account features the salary in the new position on one side and the salary at the old job on the other.[5] In contrast, Gossen's Second Law solely acknowledges comparisons in terms of marginal utility and opportunity costs. Therefore, if uCat is selected as a dimension of comparison, we will refer to this specific type of uCom as "Gossen's Accounting". Would the CEO apply this decision rule, then categorical evaluations of both jobs would have to be juxtaposed in the account.[6] If uCat and uCom interact in this fashion, the predictions of the Dual Utility Model match those of standard economic theorizing. In principle, com-

[5]To some degree, such an accounting implies that the work is identical in both positions and may thus be "canceled out".

[6]Because opportunity costs always refer to the second best option, the example implies that the current position is evaluated accordingly.

parisons along this dimension appear possible. Even more, through the algebraic glasses of constrained utility maximization, comparing marginal utility and opportunity costs is necessary. However, from a psychological perspective, they are neither necessary nor are they the only dimension to compare.

To be sure, opportunity costs may be rather salient standards of comparison if two options are mutually exclusive (e.g. binary choice settings), but in the majority of daily-life decisions goods are purchased with money which does not immediately suggest a commensurable standard of comparison.[7] Nevertheless, according to economic theorizing, marginal utility effortlessly maps into the monetary dimension as a *reservation price* p_{res} (e.g. Amir, Ariely, & Carmon, 2008). Therefore, if p_{res} is compared with the monetary price p_{sale}, Gossen's Accounting would be transformed into an arithmetic operation. If the theory would be interpreted literally, a homo oeconomicus walking through the aisles of a supermarket will simply compare the p_{res} and p_{sale} of each product and purchase if the former is higher than the latter.[8] From a psychological vantage point, representing and storing such complex evaluations like marginal utility in a monetary format seems unlikely. In contrast, however, consumers may often know the range of prices for a certain product and perform comparisons in the monetary dimension. Alternatively, a seller may provide a standard in the context of a sale. For example, a reduction of the price (in percent) elicits a very easy comparison with a standard in

[7] Certainly, it may be assumed that p_{sale} reflects the opportunity cost of a purchase in terms of the consumptive potential. However, this a strong assumption regarding the psychological reality.

[8] Simultaneously, the difference $p_{res} - p_{sale}$ is a monetary measure of the advantage a consumer takes in a transaction, i.e. the consumer surplus (see Slesnick, 2008).

the past. Also, a selling price may be announced to be increased after given period. Most important perhaps, a consumer's attention may be directed toward a higher price from a competitor. All these instances may be summarized under the notion of *reference prices* p_{ref} (e.g. Kalyanaram & Winer, 1995; Köszegi & Rabin, 2006; Mayhew & Winer, 1992; Winer, 1986). In more detail, external reference prices are present in the purchase environment (e.g. suggested retail prices or prices of the product from another time or location), whereas internal reference prices are stored in memory. For example, Köszegi and Rabin (2006) developed a formal model where reference prices are determined by the prices encountered in previous time periods.

In an attempt to account for different comparisons in the monetary dimension, Thaler (1985/2008) formulated a duality between "acquisition utility" and "transaction utility". That is, acquisition utility refers to Gossen's Accounting represented as $p_{res} - p_{sale}$, whereas transaction utility hinges "solely on the perceived merits of the deal" (Thaler, 1985/2008, p. 19). Specifically, a novel mental account is opened which contains $p_{ref} - p_{sale}$. Obviously, both utility components are comparative in nature and differ merely in their reference points. However, the transaction utility component additionally captures the evaluation of exchange itself. Thus, the actual transaction, i.e. the purchase of a certain good at p_{sale}, is compared to a reference transaction described as the purchase of the same good at p_{ref}. In order to support this theorizing, Thaler (1983) conducted a set of small experiments on hypothetical situations. One scenario involved either the purchase of a radio ($35) or television set ($650). In both cases the consumer then learned that the respective product would be $10 cheaper in another store 20 minutes

away. Not surprisingly, much more people were willing to drive 20 minutes for the radio than for the TV set. According to Gossen's Accounting alone, the discount should be evaluated in terms of its opportunity costs (i.e. time, gasoline, etc.) which is identical in both versions of the scenario. But on the transaction utility account, comparing prices of the radio suggested a good deal which translated into a high willingness to make the trip. In contrast, the same absolute discount on the TV set could not justify the extra effort.

Even more fundamental than translations of marginal utilities into reservation prices, the developments in Expected Utility Theory coined a numerical concept of utility going beyond prices (von Neumann & Morgenstern, 1944/1955). Specifically, EUT assigns cardinal values to any good or outcome which reduces every decision process to a comparison of these numbers. In principle, such comparisons are assumed to be equivalent to Gossen's Accounting despite crucial differences in the dimensions of comparison. From a psychological perspective, however, the possibility of this comparison may be met with serious doubt as complex evaluations are not likely to be represented numerically. Nonetheless, any form of this "Monetary Arithmetic" also constitutes a manifestation of uCom.

3.2 Utilitarian Dynamics

Fundamentally, categorical and comparative evaluations interact regarding the attribute on which comparisons are performed. Specifically, evaluative superiority or inferiority in some dimension of comparison presupposes a categorical evaluation of the corresponding attribute. For example, the statement that one wine is more fruity than another only implies superiority in an

evaluative sense, if the attribute "fruity" has been evaluated positively beforehand. Hence, categorical evaluations concerning the "poles" of the dimension precede comparative evaluations along that dimension. More specifically, after uCat has been inferred it becomes an attribute of a good which may serve as a dimension of comparison. A similar line of argument has been endorsed by Brunswig (1910) who proposed that comparisons of values are secondary judgments, whereas judgments about value itself are primary.

Moreover, both types of utility are assumed to contribute to behavioral decisions. That is, most choices are jointly determined by both a categorical evaluation and a relative advantage. Often, uCat and uCom are in concord such that a positive categorical evaluation of a fruity wine may be accompanied by a positive comparative evaluation due to a discount. However, there may also be a conflict between both utility components. Certainly, some "good deals" (i.e., high uCom) later turn out to be completely useless (i.e., low uCat). Related to this kind of conflict, Hsee (1999) presented participants with two different chocolates which differed in price and size, but also in shape. Then, each participant was asked two questions: one about the experience of consumption (i.e. the taste) and one about their choice. Along the lines of traditional utility theory, both questions should yield similar responses, but subjects often predicted more positive experiences with one option while choosing the other. In particular, the majority preferred the experience of eating a "lovely heart shaped" chocolate A to consuming a "disgusting cockroach shaped" chocolate B. At the same time, most subjects would have chosen the \$2 chocolate B instead of the 50 cent chocolate A. Therefore, it may be hypothesized that participants engaged in a

categorical evaluation of the shape attribute (lovely is positive, disgusting is negative) regarding the experience.[9] Instead, comparative evaluations of the price attribute ($2 > $0.5) may have dominated the final decision. Consequently, uCat and uCom have been in conflict which led response differentiation and the choice of an option that is less enjoyable (see also Hsee, Yu, Zhang, & Zhang, 2003; Hsee, Zhang, Yu, & Xi, 2003; Hsee & Hastie, 2006; Tversky & Griffin, 1991). In sum, a distinction between uCat and uCom seems particularly useful in situations of discord.

Despite their joint operation in natural situations, the two types of utility may be dissociated in controlled contexts such that a decision is dominated by either categorical or comparative utility. Our model introduces the concept of "utility focus" to describe the relative contribution of each type of utility and attempts to identify the determinants of the focus. To a considerable extent, utility focus is determined by the accessibility of information and the difficulty of each judgment.

Accessibility

Concerning uCat, the accessibility of category specifications and evaluative information plays a major role. Therefore, if a painting either cannot be categorized as "a Picasso" based on the available information or if the semantic evaluation "Picasso is good" is not available, then a categorical evaluation is unlikely. Alternatively, the brand of a product may have been pre-evaluated positively (e.g. via advertising) which would create a positive product evaluation if the brand is made acces-

[9] To some degree, this evaluation may also have involved a mental simulation of eating an oddly shaped chocolate.

sible in the purchase situation. Importantly, evaluative information may either be unavailable because it has not been stored in memory or because it is not accessible, i.e. because it does not come to mind for the judgment at hand. In a related vein, Hsee and Rottenstreich (2004) performed a set of experiments indicating a dominance of categorical evaluations if the stimulus elicited stronger affective reactions. As a consequence, evaluative information was more accessible which induced a "valuation by feeling". Furthermore, research on "affect misattribution" suggests that affect, which is not considered relevant for the judgment at hand, has less influence on evaluations (Schwarz & Clore, 1983).

As far as uCom is concerned, accessibility primarily refers to commensurable standards. That is, a Picasso painting is a less commensurable standard to evaluate an apple than an orange would be. After all, paintings and fruits share few easily accessible dimensions on which they may be compared. Hence, commensurability depends on the similarity of target and standard in terms of comparable dimensions. At this point, the logic of Gossen's Accounting reveals its weakness because it rests on the assumption that "pleasure is measurable and all pleasures are commensurable" (Edgeworth, 1881, p. 59). Moreover, the standard utility model implies that opportunity costs, i.e. the commensurable pleasures that have to be foregone, are accessible standards of comparison in every decision. While this assumption may hold under certain conditions (e.g. if the scarcity of resources is salient), decisions are often based on more accessible standards of comparison or even based solely on the evaluative implications of uCat, i.e. without any comparison at all.

A related line of research has been summarized as

"General Evaluability Theory" (Hsee & Zhang, 2010). In detail, the theory distinguishes between values referring to the "level of an objective attribute" and evaluations which are "subjective reactions" towards an object (Hsee & Zhang, 2010, p. 343). Furthermore, evaluability is defined as the degree to which evaluations are sensitive to variations in value. Most importantly, the evaluability of values is determined by the accessibility of commensurable reference information. For example, evaluating options next to each other may highlight a certain dimension on which they may be compared and thus shift utility focus towards uCom. In particular, Hsee (1996) found systematic variations in the willingness-to-pay for dictionaries depending on whether they were evaluated separately (by different participants) or jointly (by the same participant). The two relevant attributes were entries (10,000 vs. 20,000) and defects (none vs. small). Importantly, while the smaller but non-defective dictionary was evaluated more positively in the single-evaluation mode (SE), the larger but defective one was preferred in the joint-evaluation mode (JE). Hence, if no standard was available to evaluate the dictionaries along the number of entries in SE, participants probably resorted to categorical evaluations of the cosmetic condition (no defects is good, small defects are bad). However, if a compatible standard was available in JE, decisions were apparently dominated by uCom. In addition, if different standards are made accessible, judgments about uCom will vary as well. Accordingly, Hsee (1998) showed that two servings of icecream were evaluated in terms of their size relative to the capacity of the container in SE but by a comparison of their absolute size in JE.

Difficulty

To a major degree, categorical judgments vary in their difficulty according to the number of attributes that have to be considered to generate a sufficiently unambiguous categorization. That is, the degree of representativeness (see Rosch, 1973) determines the ease of categorization. To be sure, if sufficient processing time, intention and cognitive capacity is available (see Strack & Deutsch, 2004), even difficult categorizations will be performed. Alternatively, decision makers may resort to evaluative "heuristics" like using feelings as information (Schwarz, 2011).

For comparative evaluations, difficulty may refer to two different aspects. First, target and standard may be commensurable in multiple dimensions which may reveal conflicting judgments. For example, two apples can be easily compared if they differ only in size. Instead, comparing apples with oranges is more difficult because there exist many dimensions that afford comparison. Consequently, comparing the "ease of eating" may favor the apple while comparing vitamin C contents may favor the orange. Therefore, we propose that difficult comparisons often require trade-offs between different dimensions.[10] Simultaneously, comparisons along a given dimension are assumed to receive less weight if the comparative evaluation is made more difficult by providing an additional dimension (Tversky et al., 1988). The logic of this hypothesis may be illustrated by the "asymmetric dominance" paradigm introduced by Huber, Payne, and Puto (1982). Roughly, one

[10]For example, monetary gambles, which are often used as stimuli in experimental economics, are perfectly commensurable because they only have two attributes: payoff and probability e.g. the \$-bet and the P-bet in Lichtenstein and Slovic (1971). At the same time, they generally require a trade-off between these dimensions.

choice option dominates another if it is superior on every dimension. Therefore, dominance entails high commensurability and ease of comparison. For example, Ariely and Jones (2008) argued that a trip to Rome, during which breakfast coffee is not included, is dominated by the same trip where coffee is already covered. As a consequence, dominated options in general, and the trip to Rome without coffee in particular, would not be chosen if contrasted with the dominant Rome plus coffee option. However, actual decisions rarely include dominated options, but the trip to Rome may instead be compared to a trip to Paris. Here, the arguably difficult decision relies on some sort of trade-off between several dimensions (e.g. cultural life, food, etc.). However, adding the inferior, and therefore irrelevant, Rome without coffee option creates *asymmetric* dominance because the additional option would be dominated by the other trip to Rome but not by the trip to Paris. As a consequence, the "decoy" trip without coffee unobtrusively highlights a dimension of comparison (i.e. coffee supply) which suggests an advantage of the coffee-included trip. Not surprisingly, several studies showed that evaluations and preferences shift towards asymmetrically dominant options (e.g. Ratneshwar, Shocker, & Stewart, 1987; Simonson, 1989; Sedikides, Ariely, & Olsen, 1999). From our dual utility perspective, these shifts reflect a decreased difficulty of comparison and therefore an increased weight of uCom. Importantly, these findings contradict Rational Choice Theories where decisions are made by comparing marginal utility and opportunity cost. Therefore, because opportunity cost always refer to the marginal utility of the second best option (Buchanan, 2008), the decoy option must not affect decision making from the

perspective of traditional economic theorizing.[11]

The second type of difficulty does not focus on attributes and dimensions but on the cognitive operation that has to be performed. In detail, the difficulty of a comparison depends on the necessary computations. Obviously, it is easier to recognize that a stock has lost half its value when it dropped from 20 to 10 than when it dropped from 2562 to 1281. Moreover, comparative judgments may be facilitated by a certain visual presentation (e.g. pie diagrams) or by directly referring to the distance between target and standard, ideally in a relative format like a deviation in percent. For example, the nutritional inferiority of an apple is easily recognized by stating that it has only 20% of the vitamin C an orange has. In addition, the difficulty of the mental operation depends on the dimension of comparison. That is, arithmetic operations can probably be considered easier than comparisons of abstract evaluations like marginal utility and opportunity cost. After all, Gossen's Accounting not only requires that categorical evaluations have already been established, but also that these evaluations are mentally represented in a commensurable format. Certainly, it is rather difficult to evaluate a career as a scientist or as a rockstar in a format that allows an easy comparison.

In the next chapter, we will test selected predictions of the Dual Utility Model.

[11] In detail, subjects assigning higher marginal utility to the decoy than to the non-dominant option would have chosen the asymmetrically dominant option irrespective to the decoy's presence. Simultaneously, subjects assigning higher marginal utility to the non-dominant option should only consider the potentially asymmetrically dominant option in terms of opportunity cost because the decoy will never be second best under these conditions.

Chapter 4

Test of the Dual Utility Model

The last chapter described the central components of the Dual Utility Model as well as their interactions. Because the traditional approach to the concept of utility paved the way for various applications as well as for diversified body of evidence documenting its failures, the following tests of the Dual Utility approach will also take place in a variety of experimental settings. In the end, this multifaceted array of paradigms might not only demonstrate the integrative power of the proposed model but also the generalizability of its predictions.

The first study addresses a specific anomaly from the domain of prosocial behavior (Heyman & Ariely, 2004). While the currently accepted explanation centers around the construal of social relations, we suggest that systematic variations in the accessibility of commensurable standards may better account for the anomalous findings. Despite the broad acceptance for simplified paradigms among researchers from Judgment and Decision Making, the second study transfers the logic of the first study into a more native domain of

economic theorizing. Therefore, standards of comparison are introduced into an allocation decision where the traditional utility paradigm can be applied without any theoretical objections. Our third study concerns the "crowding out hypothesis" for two economically important dimensions and standards of comparison. In detail, the experiment juxtaposes Gossen's Accounting and Monetary Arithmetic as ways of constructing mental accounts and tests the implications of such multidimensional comparisons. In the fourth experiment, we address a secondary hypothesis about the circumstances when decision makers spontaneously resort to Gossen's Accounting. More specifically, as this type of comparison requires opportunity costs as a standard of comparison, we test whether increased scarcity salience induces decision makers to consider them. The last two experiments explore the utilitarian dynamics in the Ultimatum Game (Güth, Schmittberger, & Schwarze, 1982). In particular, the fifth experiment was designed to test how the difficulty of comparative computations affects responder behavior. Finally, in the sixth experiment, the consumptive potential of money was highlighted in an attempt to induce Gossen's Accounting in responders.

4.1 Comparative Utility and Prosocial Behavior

The first study investigates how the comparative evaluability of different benefits affects the willingness to exert effort in social interactions. In modern societies, where the individual level of specialization became very high, most people focus their efforts on specific tasks they have chosen as their profession (for a formal model

see Yang, 2001). At the same time, this process was accompanied by increasing degrees of the division and exchange of labor. Under the assumption that labor is not intrinsically positive, any effort must be compensated by some payment. More specifically, economic agents are assumed to work only if the marginal utility of the payment outweighs the opportunity cost of exerting the effort. Consequently, if paying more increases marginal utility, then the willingness to exert more effort will also increase.

In addition, if the division of labor in a society reaches higher levels, exchanges either rely on a coincidence of wants (Jevons, 1875; Menger, 1892) or some kind of currency system, i.e. money. In general, economic theorizing assumes that behavior is unaffected by the introduction of a medium of exchange (see e.g. Yang, 2001). That is, as long as the marginal utility of a payment is kept constant, it is irrelevant whether the exchange is a barter, i.e. payment happens in goods directly, or whether money is chosen as the mode of payment.

However, Heyman and Ariely (2004) found that the mode of payment affects the relationship between its scope and the corresponding willingness to exert effort. In detail, they propose that the type of payment determines which relational model (A. P. Fiske, 1992) is applied to the social interaction. Broadly, Fiske's framework consists of four different relational models which may be organized in two categories. First, Communal Sharing, Authority Ranking and Equality Matching all share features of close social relations (e.g. kinship) where exchanges are often restricted by strong bonds between the interacting parties. In contrast, Market Pricing, which alone constitutes the second category, refers to often anonymous exchanges between parties

whose interaction is often restricted to isolated transactions. According to Heyman and Ariely, these two groups of relational models manifest in either "monetary or social markets" characterized by monetary or non-monetary payments. Most important, the applied relational model is assumed to determine whether the behavior is conceptualized by trading-off costs and benefits (i.e. by applying mental accounting) or guided by social norms. That is, Heyman and Ariely (2004) implicitly propose that cooperation and altruism may serve as ends in themselves which are able to compensate for the effort. In several experiments the authors showed that effort corresponds to the level of payment in monetary markets. In contrast, this correspondence disappears in social markets even though the monetary value of payments was controlled for. Also, the authors interpret the finding that effort depends on the level of payment in mixed markets, where effort is paid with goods, but money is introduced by pricing these goods, as an indication of a shift in the social relations. In particular, the authors conclude that prices function as a signal to invoke the economic relational model.

In contrast, the framework proposed here states that increasing the scope of benefits will only translate into increased efforts if that increase is also reflected in the subjective evaluation of the benefits. To be sure, relational models might affect behavior but we interpret the findings of Heyman and Ariely (2004) as a lack of an accessible and commensurate standard of comparison. Probably, knowledge about potential standards (Hsee & Zhang, 2010) is available for monetary payments (e.g. the minimum wage) which ensures both, comparative evaluability and the correspondence of effort and payment. However, standards are much less likely to exist for non-monetary pay-

ments because most people are rather inexperienced in bartering their labor for goods. Interestingly, in all studies by Heyman and Ariely (2004), low levels of payment in monetary and mixed markets led to lower effort levels whereas effort in all other conditions (including a no payment condition) was higher but independent of the level of payment. For example, the effort in their first experiment is described as moving a sofa, a task which takes about 15 minutes (see our own results in the Appendix) and is paid with $0.50 in the low monetary payment condition. In contrast, participants in the low non-monetary payment condition anticipated being paid with a candy bar. Consequently, while a $2-hourly wage might be considered a "bad deal" (Thaler, 1983, 1985/2008) in comparison to the reference wage, receiving some candy bars might not elicit corresponding thoughts.[1] As a consequence, the willingness to exert effort decreases for low monetary payments but not for low non-monetary payments.

The purpose of the first study was to deconfound the "evaluability" and type of payments. Therefore, we employed the joint-single evaluation paradigm (Hsee, 1996) to make standards of comparison for non-monetary payments more accessible.[2] Hence, we predicted no correspondence between non-monetary payment level and effort in single evaluation mode. In contrast, we hypothesized that effort levels will vary with payment levels under joint evaluation even for non-monetary payments. Less surprisingly, jointly evaluated monetary payments should also create corre-

[1] Importantly, this reasoning for monetary markets also applies for "mixed market" because pricing a good gives rise to the same reference information.

[2] In the original Heyman and Ariely (2004) study every participant only evaluated one payment level, i.e. all were in the single-evaluation mode.

sponding effort levels.

Sample and design

Two-hundred US-citizens (age: $m = 33.16$, $SD = 10.64$, 45% female) were recruited via *Amazon Mechanical Turk* (mTurk) for an online study which could be performed at their own computers. We advertized a payment of $1.50 for the participation in a 15 minute study on judgment and decision making. Users who agreed to participate were redirected to an external site (*EFS Survey*) where they started the survey.

Evaluation mode (*joint vs. single*) and payment type (*money vs. good*) was varied between subjects. Therefore, the payment level (*low vs. high*) varied within subjects in the joint evaluation (JE) conditions and between subjects in the single evaluation (SE) conditions. Because this study focused mainly on the evaluation of non-monetary payments, we compared non-monetary payment in JE with monetary payment in JE to explore how different types of payments are evaluated in comparison with a standard. Therefore, we did not include monetary SE conditions where a effort-payment correspondence has already been shown (Heyman & Ariely, 2004). Thus, the design consisted of four cells: non-monetary payment (JE), monetary payment (JE), *low* non-monetary payment (SE) and *high* non-monetary payment (SE).[3]

Materials and procedure

We tested hypothetical situations instead of real behavior, however, Heyman and Ariely (2004) assert that

[3]In fact, the study design also included a no payment condition as a control group. However, the results will not be reported further because they are of very limited relevance for the question at hand.

results from intuitions about hypothetical situations can be generalized to real behavior. Thus, participants were presented with five scenarios (see Appendix) in a randomized order. Also, following Heyman and Ariely (2004), helping behavior was measured indirectly to minimize effects of social desirability. Therefore, participants were asked to rate the willingness-to-help (WTH) of an average person instead of their own. In the SE conditions, we provided one Likert scale anchored at *1 = not at all likely to help* and *11 = will help for sure*. In the JE conditions, we provided one such Likert scale for each level of payment (see Appendix).

After indicating helping efforts for all scenarios, all participants answered several questions concerning the duration (in minutes) and the strain (Likert scale: *1 = not exhausting at all* to *5 = very exhausting*) of the five tasks described in the scenarios. In addition, all participants estimated the prices of the non-monetary payments and then answered some demographic questions.

Results

Ten separate one-factorial ANOVAs of the prices for both (low and high) non-monetary payments in each scenario showed that there were no differences across cells (all $ps > 0.10$). Also, the estimated duration of all five tasks did not differ across between-subjects conditions (all $ps > 0.15$). Similarly, there were no differences between cells for the estimated strain of the different tasks (all $ps > 0.30$) (see Appendix for a more detailed analysis of the payments and tasks).

Helping

For each level of payment, willingness-to-help was averaged across all five scenarios (see Figure 4.1).

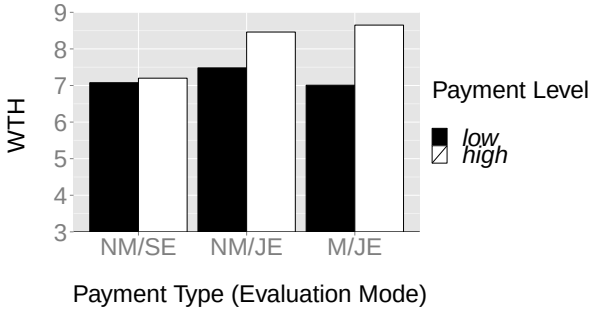


Figure 4.1: The figure shows the average willingness-to-help (WTH) in exchange for different payment levels. Participants were either offered non-monetary payment (NM) or monetary payment (M) in either separate (SE) or joint evaluation mode (JE). Note that higher payments only corresponded to higher efforts if comparative evaluations were possible.

In the SE conditions, the average WTH across all five scenarios was 7.08 for low non-monetary payments and 7.20 for high non-monetary payments. Therefore, we replicated the finding of Heyman and Ariely (2004) that the effort-payment correspondence does not necessarily hold for non-monetary payments ($t(95.45) = 0.33,\ p = 0.74$). In contrast, if non-monetary payments were evaluated jointly, the average willingness-to-help was 7.48 for low payments and 8.46 for high payments. That is, the correspondence held for non-monetary payments if they were evaluated jointly ($t(48) = 6.42, p < 0.001$).

Furthermore, a 2 (payment type, between subjects) $\times$ 2 (payment level, within subjects) ANOVA for both joint evaluation conditions showed no main effect of payment type ($F(1, 98) = 0.17,\ p = 0.68$) but a significant main effect of payment level ($F(1, 98) = 72.98,\ p < 0.001$).

Importantly, this analysis also reveals a significant interaction of payment type and level ($F(1, 98) = 4.67$, $p = 0.033$). In particular, the difference between low and high payment is larger in the monetary payment condition ($d = .90$) than in the non-monetary payment condition ($d = .50$).

Discussion

The results confirm our hypotheses that the evaluability of payments determines how much effort is exerted in exchange for payments of different scope. If no commensurable standards are accessible to evaluate payments, then different levels of payment are not reflected in corresponding effort levels. However, while this is often true for non-monetary payments in general, artificially introducing a standard of comparison restores the effort-payment correspondence. Therefore, even though the payment type might still suggest an transaction in a "social market", efforts depend on payments if they can be evaluated comparatively. Hence, transactions by themselves are only predictive of behavior if evaluative processes are taken into account. More specifically, the proposed duality between uCat and uCom may explain why high evaluability transactions become more attractive if payment is increased but those allowing no comparative evaluation do not. These findings suggest that exploring the judgmental processes that create utility might parsimoniously explain anomalous behavior without hasty appeals to factors levering out the economic approach entirely.

However, even if the evaluability of the payments is controlled in JE, the mode of payment still has influence on behavior. In detail, our results indicate that effort is more sensitive to the level of monetary pay-

ments than it is to the level of non-monetary payments. Two explanations might be possible. First, in line with the Fiske-Heyman-Ariely reasoning, this could be interpreted as a certain relevance of the payment type for the construal of social relations. In that sense, these results point towards a tendency to evaluate transactions in social markets in a categorical fashion. In fact, A. P. Fiske (1992) also proposed that Market Pricing involves comparisons "in terms of cost-benefit ratios" (p. 692), whereas the cognitive operations associated with the three additional relational models summarized as social markets are categorical in nature. Specifically, parties in Communal Sharing must belong to the same social category (e.g. family), while Authority Ranking additionally requires an hierarchical ordering of these categories. In contrast to categorizing the other person in the transaction, categorizations in Equality Matching refer to the exchanged goods. In sum, these categorical operations could have shifted utility focus away from uCom. As a consequence, comparisons would be less likely to determine behavior. However, adhering to social norms to explain anomalous behavior not only appears to be an ad hoc reaction but also would have to be excluded from the explanatory realm of economics according "instrumentality criterion" (Bruni & Sugden, 2007; Pareto, 1909/1971). Therefore, the second line of explanation might be more parsimonious. Specifically, goods might simply be less commensurable standards of comparison. Even though in this study low and high non-monetary payments always belonged to the same category (chocolate, wine, etc.), comparable dimensions might have been less accessible or the cognitive operation might have been more difficult. Therefore, utility focus would have shifted away from uCom. Unfortunately, a distinction between these

explanations based on the current findings is not possible. Hence, future investigations should systematically vary both the category membership of transaction partner (i.e. the social relation) and the good serving as a standard of comparison (both, the category and the difficulty of comparison).

To a certain degree, the paradigm used in this study posits some limitations to the interpretation of the findings. First, behavior served as an indirect measure to evaluate different payments which might have masked the value associated with alternative uses of the inevitably scarce resource time, i.e. the opportunity costs of getting paid. Moreover, this one-sided focus on the benefits of the transactions could have been intensified by demand effects due to the task's structure. Specifically, directly juxtaposing two alternative payments might induce participants to artificially differentiate their responses. To be sure, measuring only one side of an transaction is common in JDM research, but economic theory needs to be considerably reformulated and simplified in order to apply to these situations. At the same time, the introduction of reference information via a combination of different payment types and evaluation modes is rather indirect. That is, despite the crucial role different standards of comparison play in the assessment of uCom, the current paradigm lacks coherent comparisons across conditions. Furthermore, the domain of helping others is a very specific form of social interaction where norms have an influence on behavior. Therefore, it might not be an adequate basis to generalize the results. All these issues were addressed in the design of the next study where the effect of comparative utility is explored in a simple allocation decision.

4.2 Comparative Utility in Resource Allocation

In general, economic theorizing conceptualizes human decision making as the allocation of scarce resources (i.e. an endowment) to different goods (e.g. G. S. Becker, 1965; Gossen, 1854). Moreover, the composition of consumption bundles can be changed by transactions according to the prices in the market. That is, these conditions dictate which quantity of one good may be exchanged for a certain quantity of another. As has been outlined above, the use Gossen's Accounting further implies that the respective quantities involved in each transaction are compared in terms of marginal utility and opportunity costs generated by categorical evaluations. To some degree, allocation decisions probably make opportunity costs more salient than purchase decisions because changing a bundle's composition inevitably involves giving up some goods in favor of others. Consequently, the predictions of the standard utility model are ideally investigated under these circumstances.

However, beyond serving as a unit to standardize marginal utility and opportunity costs, prices may also serve as a reference information allowing another type of comparative judgment. In particular, a price may be compared to some reference price which can exert profound influence on the evaluation of a transaction (Kalyanaram & Winer, 1995; Monroe, 1973; Thaler, 1983, 1985/2008; Winer, 1986). To be sure, this Monetary Arithmetic is irrelevant according to Gossen's approach but it is also reflected in uCom.

In order to minimize the availability of existing reference prices, the allocation task in this study was set in an environment that is unfamiliar to most persons:

fish keeping.[4] Therefore, participants were given a hypothetical scenario consisting of two rounds. In each round they had to spend a certain budget on fish for a new aquarium. Two kinds of fish were available: Guppies and Tetras. But whereas the price of Guppies remained constant over the whole experiment the price for Tetras changed from the first round to the second. As a consequence, the demand for Tetras in the first round is very unlikely to be influenced by Monetary Arithmetic, whereas in the second round the prior price is accessible as a standard of comparison.

Thus, we hypothesized that if the price is higher in the second round participants would "underspend" in relation to participants who encountered the higher price in the first round. In contrast, participants confronted with the lower price in the second round would "overspend" relative to participants facing the low price in the first round. Following Thaler (1983, 1985/2008), the prices in the second round suggest a "rip-off" in the first case but a "good deal" in the latter. Consequently, we predicted an order effect such that increasing prices generate a lower overall demand than decreasing prices.

Sample and design

Five-hundert-thirty-four US-citizens (age: $m = 31.71$, $SD = 9.9$; 36% female) were recruited for an online study via Amazon mTurk like in the previous study. They received \$0.3 for a 3 minute experiment.

Participants were tested in a 2 (order: decrease vs. increase; between subjects) $\times$ 2 (price: high vs. low;

[4] I want to thank my dear friend Daniel Kemmling for suggesting this particular setting for the experiment.

Table 4.1: Experimental design of the Guppy Game. Please note that $p_1 < p_2$.

	Condition			
	Increase		Decrease	
Round	uCom	Price	Price	uCom
1	-	p_1	p_2	-
2	Rip-Off	p_2	p_1	Good Deal

within subject) mixed design illustrated in Table 4.1. The demand for Tetras served as a dependent variable.

Materials and procedure

The hypothetical scenario described a spontaneous purchase of an aquarium. In two subsequent rounds, participants had a budget of \$20 to buy Guppies (\$1) and Tetras (*low*: \$2; *high*: \$3 / \$4).[5] Therefore, only two pieces of information were available about the fish: their name and their price. Participants used two independent scales to indicate how many Guppies and how many Tetras they would buy.

Participants read the scenario and then indicated their first allocation. The scenario continued and participants were asked to imagine that after enjoying the first aquarium so much they decided to purchase another one for a different room and indicated their second allocation with the other prices. Afterwards, participants indicated their fish keeping expertise, price estimates for both fish and answered some demographic questions.

[5]Two different prices were used in the high-condition. However, for the question at hand the differences are irrelevant and will not be reported further.

Results

Twenty-six participants were excluded from the analysis because they spend more than their available budget which left a total sample size of five-hundred-and-eight participants. The demand for Tetras across conditions is illustrated in Figure 4.2.

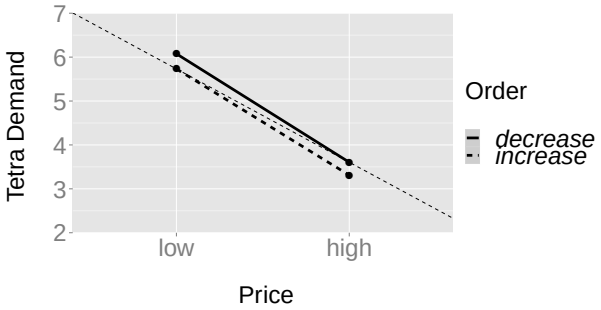


Figure 4.2: The figure shows the demand for Tetras at different price levels. Participants indicated their demand in two following rounds while the order of price presentation was either decreasing or increasing. The black dotted line represents the baseline-demand in round 1. At each price level it can be seen that the demand in round 2 deviates from demand in round 1 generating the predicted order effect.

A 2 (order: decrease vs. increase) $\times$ 2 (price: low vs. high) mixed model ANOVA yielded a significant main effect of price ($F(1, 506) = 1000.98,\ p < .001$) and a significant effect of order ($F(1, 506) = 6.37,\ p = .011$). However, in contrast to the last study, this effect of the reference price was rather small ($d = 0.22$). No significant interaction was found ($F(1, 506) = 0.10,\ p = 0.75$). In addition, planned contrasts indicate that participants who encountered prices in a decreasing order demanded on average 0.35 Tetras more in the low price condition ($t(505.92) = 2.03,\ p = .043$). Simultaneously, participants who encountered prices in an increasing order demanded on average 0.26 Tetras less in the high price

condition ($t(499.70) = 2.40,\ p = .017$).

Discussion

The results clearly indicate that higher prices caused a decrease in demand. Obviously, this is in line with the fundamental theorem of traditional economic theory "that the demand curve for any commodity, real income held constant, must be negatively inclined" (G. S. Becker, 1962, p. 3). However, G. S. Becker (1962) also argued that any increase in price must almost necessarily lead to a decrease in demand because higher prices simply exclude certain persons who cannot afford to purchase the good anymore: "[e]ven irrational decision units must accept reality and could not, for example, maintain a choice that was no longer within their opportunity set. [...] Systematic responses might be expected, therefore, with a wide variety of decision rules, including much irrational behavior" (p. 12). That is, with the maximum number of Tetras available in the budget being lower if prices were high, a decrease in average demand is ceteris paribus necessary.

More importantly, the evidence clearly shows that participants encountering decreasing prices overspend in the second round relative to participants who encountered the low price in the round before. In contrast, participants confronted with increasing prices underspend in the second round relative to participants who faced the high price before. Therefore, the differences implied in the terms over- and underspending are manifestations of additional uCom due to Monetary Arithmetic between the sale price and the reference price from the prior round. Note that the results are also completely in line with Transaction Utility Theory (Thaler, 1983, 1985/2008). Furthermore, the order

effect suggests that preference discovery is not path independent (see Bruni & Sugden, 2007). Therefore, the evidence rather supports the Preference Construction Hypothesis (Slovic, 1995) than the Preference Discovery Hypothesis (e.g. Plott, 1996).

In addition, these results support the notion of two distinct evaluative processes. Specifically, even though the unfamiliarity with the subject renders price comparisons in the first round unlikely, everyone was obviously able to evaluate both fish. For example, participants may have based their evaluation on a categorical judgment about the fish's names. To be sure, we do not claim that comparative judgments do not play a role in the first round as directly comparing the prices of both options may elicit inferences about quality (Dodds, Monroe, & Grewal, 1991; Rao & Monroe, 1989) or prestige (Bagwell & Bernheim, 1996; Braun & Wicklund, 1989). However, these inferences still apply in the second round but are accompanied by an additional uCom component. This comparative utility is negative for participants in the increasing condition thus leading to underspending and it is positive for participants in the decreasing condition which leads to overspending.

Interestingly, the non-significant results concerning the interaction suggest that there were no asymmetries concerning the additional uCom component regarding the direction of the price change from round one to round two. Nonetheless, such an asymmetry is suggested by Prospect Theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) whereafter "losses loom larger". Thus, a price increase should cause stronger uCom effect than a price decrease. However, the results do not confirm this hypothesis.

To be sure, the effect of the reference price was quite

small in this study. However, opportunity costs might have been very salient standards of comparison in the allocation scenario. Given that all income is used, increasing the demand for one kind of fish inevitably required to forego some units of the other.[6] If a categorical evaluation of these opportunity costs provided an additional dimension, the corresponding comparison in terms of uCat could have superseded the Monetary Arithmetic. Hence, the paradigm employed in the next study features decisions about ordinary consumer products in a monetary payment system. That is, purchasing a good with money does not involve a direct decision between two options (Zauberman & Lynch Jr, 2005). Thus, in simple purchase decisions, categorical evaluations of opportunity cost are probably less accessible standards of comparison than in decisions about the allocation of scarce resources. In addition, the virtue of the "Guppy Game", which relied on unfamiliar goods to avoid the interference of existing reference prices, may also be a vice concerning the ecological validity of the findings. Thus, the next study uses an alternative paradigm in which purchase decisions about more familiar goods were investigated while the reference prices themselves and as well as opportunity cost salience were varied systematically.

4.3 Multi-Dimensional Comparisons

Undoubtedly, the preceding investigation indicated that comparative evaluations play a crucial role in both, economic theorizing and in empirical investigations of decision making. However, comparisons appear to be made on different dimensions. For example,

[6] Moreover, this effect was probably increased by presenting the response dimensions right next to each other.

the last study showed that comparing prices does have an influence on individual demand. Simultaneously, we hypothesized that uCat might also have been a relevant dimension of comparison. Unfortunately, the experimental paradigm did not allow to draw causal inferences about the interaction of different comparisons because opportunity cost salience was not systematically manipulated in the allocation scenario. Hence, the current study explores how varying the presence of different standards of comparison affects the evaluation of familiar consumer products.

Despite their prominent place in economic theory, opportunity costs seem to play a rather subordinate role in actual decision making. In fact, previous research in the domain of accounting suggests that opportunity costs are ignored in investment decisions (S. W. Becker, Ronen, & Sorter, 1974; L. A. Friedman & Neumann, 1980; Hoskin, 1983; Neumann & Friedman, 1978). In addition, psychological research found evidence for opportunity cost neglect across a broader range of more familiar situations (Frederick, Novemsky, Wang, Dhar, & Nowlis, 2009; Larrick, Morgan, & Nisbett, 1990; Larrick, Nisbett, & Morgan, 1993; Spiller, 2011). Therefore, some of the authors argue that "[p]eople will often perform a service for themselves (for example, mowing the lawn) even though the amount they could earn performing some other activity would be higher than the amount they would have to pay for someone else to perform the service." (Larrick et al., 1990, p. 363). However, Larrick and colleagues assert that training participants to consider opportunity costs reconciled their behavior with economic theory. More specifically, the training pointed out the flawed reasoning in problems similar to the lawn mowing situation quoted above. Thus, despite not being a a very

accessible standard of comparison, it is possible to induce subjects to consider opportunity costs in evaluative judgments. Nonetheless, the procedure employed as a training does not clarify the exact psychological mechanisms involved in these evaluations.

One possible mechanism involved in the judgment of opportunity costs was investigated by Shiv and Huber (2000). In detail, they proposed that consumers construct mental images to simulate the consumption experience. Therefore, goods could be evaluated on the basis of the simulated experience. In a related vein, Kahneman and Miller (1986) argued "that events are sometimes compared to counterfactual alternatives" (p. 136). That is, based on the context of the decision, scenarios about non-existent situations are constructed which then serve as a standard of comparison for the factual event. As a consequence, providing certain cues stimulating the mental imagery associated with alternative resource allocation might make opportunity costs more relevant for actual decision making.

In contrast, comparing prices to evaluate a transaction only requires simple arithmetic operations. Often, monetary standards of comparison are classified into external or internal reference prices (Mayhew & Winer, 1992). However, sometimes internal reference prices may not merely be recalled but result from a judgment. Most prominently, Tversky and Kahneman (1974) have shown that judging whether some unknown target value is lower or higher than an arbitrary standard (the anchor) biased the subsequent absolute judgment of the target value. An explanation for this "anchoring effect" was offered by Strack and Mussweiler (1997) who showed that the anchor selectively increased the accessibility of information of anchor-consistent information (a more elaborated model was

presented in Mussweiler and Strack (1999)). For example, Mussweiler, Strack, and Pfeiffer (2000) proposed that contemplating whether the average German car costs more or less than 60.000 Euros activates knowledge about luxury cars, whereas asking whether it costs more or less than 10.000 Euros activates knowledge about smaller cars. The absolute judgment of an average car's price then draws on the selectively activated knowledge. That is, high estimates are evoked if knowledge about luxury cars was made accessible by an high anchor while low estimates result if low anchors made knowledge about small cars more accessible.

In this study, we used the anchoring effect to manipulate participants internal reference prices. Similarly, Ariely, Loewenstein, and Prelec (2003) found that arbitrary anchors affected the bids participants placed on consumer products (e.g. a high class wine). More important however, the bids for subsequently offered products from the same category (e.g. a wine of lower quality) proved to be coherent. That is, despite being arbitrarily biased by the anchors, the bids for the high quality wine still served as a standard of comparison to make the bids for the lower quality wine. Therefore, we hypothesized that presenting a high anchor for the reference price estimates leads to a more positive evaluation of the transaction because the actual sale price is compared to a high reference price. By the same logic, low anchors should lead to less positive transaction evaluations. More specifically, we assumed that the effect of the anchor on the evaluative judgment is mediated by the arithmetic difference between the price estimate (i.e. the reference price) and the actual sale price. Note that the procedure guarantees that only this comparative utility component is altered while ac-

tual costs constant remain constant. In addition, we varied the salience of opportunity costs following the distinction implied in Becker's model of time allocation. Specifically, G. S. Becker (1965) suggested that utility may stem from leisure time and the consumption of goods. Consequently, we implemented two different foci for the counterfactual simulations directed either towards alternative consumption opportunities or on the reduced requirement to work whereby more time for leisure is made available. Most important, we hypothesized that if opportunity costs are made salient as potential standards of comparison, the effect of the price comparison on the evaluation of the transaction will decrease.

Sample and design

We recruited one-hundred-eighty-one US-citizens for an online survey (about 15 minutes) via Amazon mTurk.[7] The advertised compensation was $1.50.

Participants were tested in a 3 (opportunity cost cue: none vs. time vs. consumption, between subjects) $\times$ 2 (anchor: low vs. high, within subject) mixed design. At the beginning, participants were randomly assigned to one of the opportunity cost conditions. The anchors in the evaluation task were randomly drawn for each trial. Dependent variables were price estimates, purchase intentions and an affective evaluation of the transaction. The data structure is hierarchical in this setting because the different anchors are nested inside participants which are assigned to different opportunity cost salience conditions.

[7] Due to a data recording problem demographic information is not available for this study.

Materials and procedure

On the first page of the questionnaire, subjects read a short introduction about a hypothetical trip to the shopping mall where several products caught their attention. In total, the evaluation task included twelve products which were presented in a randomized order. For each product, participants were first given a short description of the item (see Appendix for an overview over all stimuli) which remained on the screen for all subsequent questions. On the next screen participants were asked to indicate whether the current product was more or less expensive than an anchor X. The anchor was either low or high for each item whereby the pairings of products and anchors were randomized.

After the relative judgment about the price, participants estimated the absolute price of the product (from now on the reference price p_{ref}) before learning the actual sale price p_{sale} on the subsequent screen (see Appendix). According to the experimental condition, the sale price was presented either without or with an additional opportunity cost reminder. In the time condition, participants were shown the cue: "*At an hourly wage of \$10 per hour you would have to work Y additional hours to afford it.*" The value Y equaled p_{sale} divided by the hourly wage of \$10. In contrast, participants in the consumption condition read the following cue: "*Assuming an average price of \$10 per pizza you could eat Z pizzas instead of buying it.*" In this case, the value Z equaled p_{sale} divided by the assumed price of a pizza.[8]

[8]For each opportunity cost condition there were two versions of the additional information which were randomly selected for each product. Specifically, the hourly wage in the time condition was either \$10 or \$20 and the alternative good was either pizza or a visit to the movie theater (both assumed to cost \$10) in the good condition. However, problems with the programming prevented the

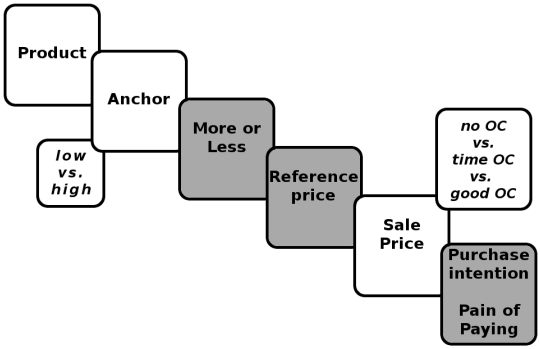


Figure 4.3: The procedure depicted in this figure was repeated for each of the twelve products. The anchor for each product was randomly selected. Grey boxes indicate measures. Depending on the opportunity cost condition, the sale price was accompanied either by no opportunity costs (no OC) or by opportunity costs expressed either in time units (time OC) or alternative consumption goods (good OC).

In the control condition, no additional information was provided.

Next, purchase intentions were measured on a scale from 1 (*very unlikely*) to 5 (*for sure*). Furthermore, the affective evaluation of the transaction was measured with a pictorial scale showing faces with varying degrees of frowning to smiling (Thomas, Desai, & Seenivasan, 2011). Figure 4.3 illustrates the sequential structure of the paradigm.

Results

We calculated the relative deviation of the sale price from the estimated reference price $\Delta p = \frac{p_{sale} - p_{ref}}{p_{sale}}$. This transformation standardizes the reference price estimates across products and preserves the intuition that a lower sale price results in a more positive evaluation

recording of the necessary data.

of the transaction.[9] Note that values for Δp larger or smaller than 400%, i.e. where p_{ref} was at least four times larger or smaller than p_{sale}, were excluded from the analysis.

Three different analyses are informative in this experimental setting. First, the anchoring effects on the price estimates, and therefore Δp, was analyzed as manipulation check. Second, a mediation analysis explored the hypothesis that anchors influence the evaluation the transaction via the the price difference Δp. Finally, we tested the moderating role of opportunity cost consideration concerning the effect of relative price differences on the transaction evaluation.

Anchoring effect on reference prices

Averaged across all products, the anchor had the predicted effect on price estimates Δp (for a more detailed analysis see Appendix). In particular, the average Δp was 0.18 in the low anchor condition and -0.32 in the high anchor condition ($t(180) = 15.27,\ p < 0.001$). That is, low anchors led to reference prices 18% below sale prices, while high anchors led to reference prices 32% above sale prices.

Relative price differences as a mediator

We tested for mediation following the procedure outlined in Baron and Kenny (1986). Therefore, the data was aggregated across low anchor products and high anchor products for each participant. To simplify the presentation of the results, we calculated a purchase evaluation index from purchase intention and affective

[9] Δp can be interpreted in the same way as the sale price p_{sale} itself because $\frac{\delta \Delta p}{\delta p_{sale}} > 0$.

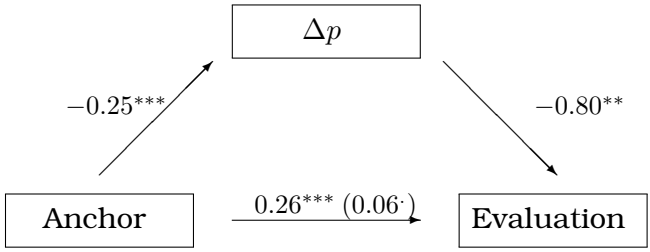


Figure 4.4: The mediation analysis shows that anchors affected transaction evaluations by altering the relative price difference between sale price and reference price $\Delta p = \frac{p_{sale} - p_{ref}}{p_{sale}}$.

evaluation (Pearson's $r = 0.79$, $t(360) = 24.42$, $p < 0.001$) which served as our main dependent variable. The independent variable was the anchor and the potential mediator was Δp. For all three regressions, parameters for multilevel models (including a constant and the respective predictor) were estimated using the maximum-likelihood procedure. In a first step, we regressed the evaluation index on a constant and a dummy variable for the anchors (*low anchor* = -1; *low anchor* = +1). The anchor had a significant effect on product evaluation ($\beta_{anchor} = 0.26, p < 0.001$). Furthermore, a second regression showed a significant effect of the anchors on the potential mediator Δp ($\beta_{anchor} = -0.25$, $p < 0.001$). Most important, in a third regression, where the evaluation index is predicted by both the anchor and Δp, the anchor only had a marginally significant effect ($\beta_{anchor} = 0.06$, $p = 0.096$) whereas the mediator Δp significantly predicted purchase evaluation ($\beta_{\Delta p} = -0.80$, $p < 0.001$). The results of the mediation analysis are illustrated in Figure 4.4.

Even though β_{anchor} from the third regression was not significant at traditional levels, the results suggest no perfect mediation, i.e. the anchor probably affected the evaluation beyond Δp. However, the ef-

fect of the anchor was drastically reduced to 23% of its original size by controlling for Δp. Therefore, the moderation analysis performed below does not include the anchors as a experimentally varied variable but the relative price differences Δp.

Opportunity cost consideration as a moderator

The moderation hypothesis was tested by fitting a multilevel linear regression model to the data. Specifically, we estimated the parameters for the following model:

$$I = \beta_0 + \beta_1 * D_{consumption} + \beta_2 * D_{time} + \Delta p * (\beta_3 + \beta_4 * D_{consumption} + \beta_5 * D_{time}) + Res. \quad (4.1)$$

The dependent variable I stands for the purchase evaluation index, while $D_{consumption}$ and D_{time} are dummy variables coding the opportunity cost conditions, respectively. As before Δp is a numeric variable denoting the relative difference between the sale price and the reference price.[10] As a consequence, the model yields $j = 3$ different functions $I = f_j(\Delta p)$ (one for each opportunity cost salience condition) describing a linear relationship between the evaluation of the transaction and the relative difference between sale price and reference price.

Table 4.2 summarizes the maximum-likelihood estimates for the regression model. As can be seen in Equation 4.1, the model assumes linear functions $f_j(\Delta p)$ which are plotted in Figure 4.5. As predicted, the parameter β_3, which denotes the slope of $f_j(\Delta p)$ in the control condition, was highly significant and negative. Therefore, smaller values of Δp elicited more pos-

[10]The constant (β_0) is included for statistical reasons but will not be reported or interpreted further.

Variable	i	β_i	*SD*	*df*	*p*-Value
Intercept	0	2.74	0.070	178	< 0.001
$D_{consumption}$	1	-0.14	0.109	178	0.208
D_{time}	2	-0.12	0.109	178	0.286
Δp	3	-1.12	0.108	178	< 0.001
$D_{consumption} \times \Delta p$	4	0.19	0.174	178	0.279
$D_{time} \times \Delta p$	5	0.43	0.155	178	0.006

Table 4.2: Estimates for the regression model described by Equation 4.1.

itive evaluations. In addition, neither β_1 nor β_2 was significant which indicates that the intercepts of the functions $f_j(\Delta p)$ do not differ if opportunity costs are made salient. More important, even though the β_4 estimate was positive as predicted, it failed to reach significance. Therefore, the slope of $f_j(\Delta p)$ is not significantly steeper in the control condition than in the opportunity costs of consumption condition. In contrast, the β_5 estimate was significant and also positive. As a consequence, the slope of $f_j(\Delta p)$ is significantly steeper in the control condition than in the opportunity costs of time condition.

Discussion

The results of this study show that comparisons on different dimensions may drive the evaluation of transactions featuring familiar products from daily life. First, we replicated the effect found in the last studies such that the evaluation of transactions is affected by rather arbitrary standards of comparison. Specifically, sys-

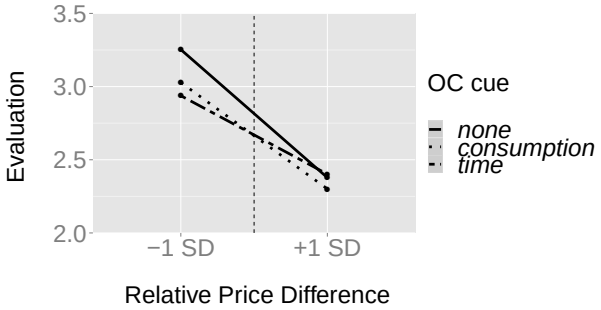


Figure 4.5: The figure shows functions $f_j(\Delta p)$ predicted from the regression model from Equation 4.1. Specifically, good deals ($\Delta p = -SD_{\Delta p}$) led to better evaluations than bad deals ($\Delta p = +SD_{\Delta p}$). The dashed vertical line marks $\Delta p = 0$. Moreover, the difference between good and bad deals was decreased if opportunity costs (OC) of time are cued. The same holds true descriptively for the consumption cue, however, this difference did not proof statistically significant.

tematically inducing high reference prices led to a more positive evaluation of the same transaction than the induction of a low reference price. Additionally, the evaluation directly hinged on the arithmetic difference between sale prices and reference prices. Hence, this finding further supports the hypothesis that the psychological process employed to generate this specific comparative utility component involves simple mathematical operations.

Furthermore, this study investigated a second type of comparison which was hypothesized to interfere with simple price comparisons. In detail, despite not having any direct effect on transaction evaluations, considering the opportunity costs of each purchase attenuated the effect of the reference price. Therefore, introducing a second dimension of comparison partly replaced price comparisons as a basis to form an evaluative judgment. More specifically, however, a certain asymmetry between different domains of opportunity costs emerged.

In particular, drawing the attention towards alternative ways to spend the time did decrease the effect of the reference price as we predicted. Yet, directing subjects' attention towards alternative consumption opportunities did not have the predicted effect. To a certain degree, this asymmetry in the results may be attributed to diminishing marginal utility itself (see also Spiller, 2011, p. 605). Thus, the opportunity costs from labor were not exactly specified, leaving open how the time not spent to earn money may be used. Therefore, decision makers might have considered a broad range of activities, potentially even varying across decisions. In contrast, the opportunity costs from consumption were relatively concrete and mentioned the consumption of specific goods. As consequence, satiation effects might have been stronger for alternative consumption. Even if we abstract from the cumulative effect due to the repeated purchase decisions, especially more expensive products necessarily involved consuming large quantities of the alternative good. For example, if it was brought to attention that instead of purchasing a TV for $750, one might also go to the movies - 75 times! Probably, working 75 hours less does not underlie the same satiation effects as these 75 visits to the cinema. Furthermore, alternative consumption might induce more thoughts about its own opportunity costs as few persons have the time to watch 75 movies. To a certain degree, opportunity costs from alternative consumption also were more specific which might have induced more "choice overload" (Iyengar & Lepper, 2000; Markus & Schwartz, 2010; Scheibehenne, Greifeneder, & Todd, 2010). Consequently, participants still had to rely on price comparisons to make the decision. However, two options do probably not constitute an excessively large choice set.

In a different vein, several researchers have proposed that consumers use the price of a product to judge its quality (e.g. Monroe, 1973; Rao & Monroe, 1989; Shiv, Carmon, & Ariely, 2005). Therefore, the main effect of Δp could be interpreted such that the reference price served as a proxy for product quality. Also, this interpretation is supported by the explanation for the anchoring effect put forward by Strack and Mussweiler (1997). Specifically, high anchors should increase the salience of information about premium products whereas low anchors should make product information connected to low budget versions of the product salient. However, this price-quality explanation fails to account for the interaction between price- and opportunity cost comparisons. After all, this would require that the validity of price as a quality cue depends on the consideration of opportunity costs.

According to Thaler (1985/2008), comparing the sale price with a reference price yields an evaluation of the transaction itself. That is, the higher the price for a given product, the worse the transaction. To a large degree, "striking a bargain" does not imply anything about the value of the goods exchanged in a transaction. Instead, it is solely a statement about advantageous terms of trade. Furthermore, if these terms are represented as mental accounts, then increasing the transparency of the transaction, i.e. the salience of costs and benefits should increase the importance of this type of advantage. At the same time, increasing the salience of costs could also direct the decision makers attention towards the scarcity of resources. As a consequence, opportunity costs might be considered, which are themselves a mere consequence of scarcity. Hence, the results from the current study support the opposite prediction that more transparent transactions, where

paying itself is highlighted, decrease the importance of advantages in terms of reference prices and sale prices. Therefore, the following experiment tested the effect of transaction transparency on the impact of Monetary Arithmetic.

4.4 Transaction Transparency

To a certain degree, research on mental accounting (Thaler, 1999) has paid more attention to the liability side of the balance sheet. Most prominently, Prelec and Loewenstein (1998) suggested that "thoughts about payment can undermine the pleasures of consumption" (p. 8). That is, transactions are evaluated differently depending on the salience of costs. Therefore, despite the traditional economic assumption that immediate consumption is preferred to delayed consumption, while delayed payments are preferred to immediate payments, several experiments documented inconsistent intertemporal preferences regarding the time of payment. Specifically, Prelec and Loewenstein (1998) found that participants preferred pre-payment for a vacation to ensure unimpaired consumptive pleasure afterwards. Therefore, the authors concluded that decoupling both sides of the mental account yields more pleasurable consumption experiences because they are not diminished by the "Pain of Paying". In fact, neuro economic investigations have shown that presenting price information about a product activates the same circuits in the brain as physical pain (see Knutson, Rick, Wimmer, Prelec, & Loewenstein, 2007).

Furthermore, subsequent investigations focused on how the mode of payment affects the pain of paying. For instance, Raghubir and Srivastava (2008) proposed that in addition to differences in temporal decoupling,

payment modes may also differ in their form (barter, cash payments, credit cards, etc.). Importantly, the form is assumed to determine the transparency of the transaction, i.e. the "vividness with which individuals can feel the outflow of money" (Raghubir & Srivastava, 2008, p. 214). More specifically, the authors compared the willingness to spend one dollar on a candy bar when participants were either endowed with cash or with a gift certificate. Supporting the transparency hypothesis, subjects were more likely to purchase with the gift certificate which was assumed to elicit less pain of paying. In a related vein, Thomas et al. (2011) investigated which influence the mode of payment exerts on "consumers' ability to control their impulsive urges" (p. 126). Therefore, the authors argue that the type of product as well as the mode of payment have an influence on the affective scenery surrounding a transaction. Obviously, certain products (e.g. high calorie snacks) are associated with positive affective reactions which may ultimately lead to impulsive transactions. In addition however, more transparent modes of payment, e.g. cash compared to credit cards, are assumed to elicit a stronger counteracting negative affective reaction. In line with their prediction, positive affective reactions towards "vice products" were canceled out by the pain of paying if the mode of payment opened a transparent balance sheet in decision makers' mental accounting system.

In sum, this line of research suggests that as transactions become more transparent, increased pain of paying leads to systematic differences in the evaluation of the transaction. To be sure, as we have already outlined in the first study, if the payment mode does not affect opportunity costs, more traditional economic theorizing deems it irrelevant for the evaluation of the

transaction. However, if more transparent payment modes direct the focus toward the payment side, then price comparisons might exert a stronger influence on the evaluation of the transaction. Specifically, instead of a pain associated with paying, making a good deal in comparison with a reference price could also elicit a certain "joy of paying". On the other hand, transparent transactions which highlight the outflow of resources may also increase the salience of scarcity. Hence, opportunity costs may become accessible as standards of comparison for the benefits of the purchase. Therefore, comparisons in the monetary dimension should have a smaller effect. These conflicting hypotheses are tested in this experiment. In addition, the study is designed as a replication of the reference price effect found in the previous study.

Sample and design

Like in all previous studies, we used Amazon mTurk to recruit one-hundred-twenty-two US-citizens (age: $m = 33.68$, $SD = 10.16$, 48% female) for an online study (15 minutes) compensated with $1.50.

Participants were tested in a 2 (payment mode: credit vs. cash, between subjects) $\times$ 2 (anchor: low vs. high, within subjects) mixed design. Parallel to the previous study, dependent variables were price estimates, purchase intentions as well as the affective evaluation referred to as pain of paying (see Thomas et al., 2011).

Materials and Procedure

The procedure was very similar to the previous study except that different product descriptions, anchors and prices were used (see Appendix). In particular, subjects

were first given a short product description, then determined whether the product was more or less expensive than an anchor price and then estimated the absolute price. As can be seen in Figure 4.6, the procedure was almost identical to the previous study. The main difference, however, was the presentation of the sale price.

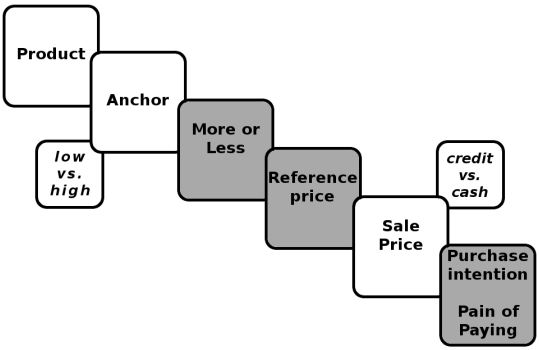


Figure 4.6: The procedure depicted in this figure was repeated for each of the twelve products. The anchor for each product was randomly selected. Grey boxes indicate measures. Depending on the payment mode condition, the sale price could either be paid by credit card or by cash.

In detail, the modified paradigm either required that participants selected which credit card to use for the purchase or what denomination of the payments they preferred, i.e. they had to select which bills to use for payment. In both experimental conditions, subjects selected one of three options which differed either in the number of bills and their denominations or in the provider of the credit card (see Appendix). Importantly, all participants were instructed to assume that every option would be available to them. After the evaluation task subjects were asked some demographic questions and then thanked for their participation.

Results

Not surprisingly, the analysis follows the same logic as in the previous experiment. In the first two steps, we confirmed the effect of the anchors on Δp and the mediating role of the relative price differences on purchase evaluations. Finally, we tested for moderation of the the latter effect by the mode of payment.

Anchoring effect on reference prices

Averaged across all products, the anchors had the predicted effect on Δp (for a more detailed analysis see Appendix). In detail, the relative deviation of the sale price from the estimated price was 0.28 in the low anchor condition and -0.14 in the high anchor condition ($t(120) = 15.81, p < 0.001$). That is, low anchors led to reference prices 28% below sale prices while high anchors led to reference prices 14% above sale prices.

Relative price differences as a mediator

The mediation hypothesis was again tested following the procedure proposed by Baron and Kenny (1986). As illustrated by Figure 4.7, the independent variable was the dummy coded anchor (*low anchor* = -1; *high anchor* = +1), the potential mediator was Δp and the dependent variable was the purchase evaluation index. The index was again calculated from purchase intentions and reverse coded pain of paying (Pearson's $r = 0.78$, $t(240) = 19.30$, $p < 0.001$). The maximum likelihood estimate from a multilevel model confirmed a significant effect of the anchors on the evaluation index ($\beta_{anchor} = 0.17$, $p < 0.001$). Second, another multilevel regression model of Δp on a constant and the dummy variable for the anchors confirmed the effect of the in-

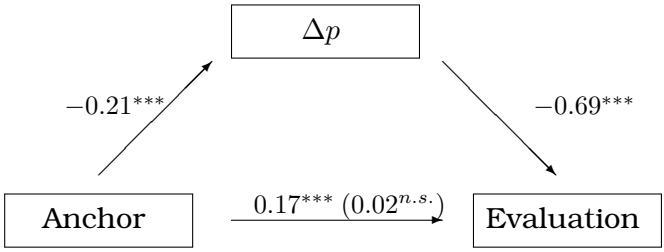


Figure 4.7: The mediation analysis shows that anchors affected transaction evaluations solely via the relative price differences $\Delta p = \frac{p_{sale} - p_{ref}}{p_{sale}}$.

dependent variable on the potential mediator ($\beta_{anchor} = -0.21,\ p < 0.001$). Finally, a third multilevel model regressed the evaluation index on a constant, the anchor dummy and Δp. The maximum likelihood estimates form this model confirmed the predicted mediation process ($\beta_{anchor} = 0.02,\ p = 0.62;\ \beta_{\Delta p} = -0.69,\ p = 0.001$).

Thus, the results suggest a perfect mediation. Specifically, the anchors did affect the product evaluation, but this effect vanished if the model controlled for Δp. Therefore, our hypothesis that the relative deviation of the sale price from the reference price affects purchase evaluation was confirmed by these results. Consequently, the anchor will again be omitted for further analyses.

Payment mode as a moderator of utility focus

In order to test whether the effect of the relative price difference on the evaluation is moderated by the mode of payment we fitted the following multilevel linear mixed-effect model to the data:

$$\begin{aligned} I &= \beta_0 + \beta_1 * D_{cash} \\ &+ \Delta p * (\beta_2 + \beta_3 * D_{cash}) + Res. \end{aligned} \tag{4.2}$$

Parallel to the model formalized by Equation 4.1 the dependent variable I denotes the evaluation index which is predicted by Δp and a new dummy variable for the cash condition D_{cash} (*cash* = 1, *credit* = 0). Due to the design modifications, this model yields $j = 2$ different functions $I = f_j(\Delta p)$ (one for each payment mode condition) formalizing a linear relationship between transaction evaluation and the relative difference.

Variable	i	β_i	*SD*	*df*	*p*-Value
Intercept	0	2.73	0.070	120	< 0.001
D_{cash}	1	-0.04	0.101	120	0.728
Δp	2	-0.57	0.146	118	< 0.001
D_{cash} x Δp	3	-0.38	0.213	118	0.080

Table 4.3: Estimates for the regression model described by Equation 4.2.

Table 4.3 shows the maximum likelihood parameter estimates for the regression model formalized by Equation 4.2.[11] In addition, the fitted values for the evaluation index from the linear model are plotted in Figure 4.8. As predicted, the basic slope parameter β_2 of $f_j(\Delta p)$ was highly significant and negative. Therefore, smaller values of Δp elicited more positive evaluations.

Furthermore, the parameter β_1 was not significant, which indicates that cash payments did not necessarily elicit more pain of paying. In contrast, β_3 is marginally significant and also negative. Furthermore, the linear

[11] Because the pairing of product and anchor was randomized one subject was presented only low anchors and one subject was presented only high anchors. Thus the degrees of freedom were lower for β_2 and β_3.

relationship describing transaction evaluation is given by $\beta_0 + \beta_2 * \Delta p$ in the credit card condition and by either $\beta_0 + \beta_1 + (\beta_2 + \beta_3) * \Delta p$ in the cash condition. Hence, the results indicate that the intercept is identical for both payment modes, but that the slope is steeper for cash payments. That is, an increase of Δp by 100% led to a decrease in evaluation of 0.57 for products paid by credit card, whereas a 100% increase of Δp led to a 0.95 decrease in evaluation of products paid in cash. Therefore, the effect of Δp on product evaluation was 67% stronger for cash purchases.

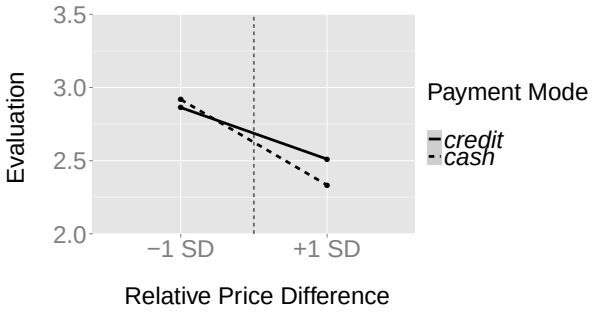


Figure 4.8: The figure shows the functions $f_j(\Delta p)$ predicted from the regression model following Equation 4.2. Similar to the last study, good deals ($\Delta p = -SD_{\Delta p}$) led to better evaluations than bad deals ($\Delta p = +SD_{\Delta p}$). The dashed line in the middle marks $\Delta p = 0$. Moreover, the difference between good and bad deals was increased if payment had to be made in cash.

Discussion

First of all, the current findings replicated the previous study which suggests that Monetary Arithmetic reliably affects transaction evaluations. That is, including a uCom component, hinging on the difference between sale prices and potentially arbitrary reference prices, crucially improves the explanatory power of the utility

paradigm. Also, the present results further levered out an explanation of more positive evaluations based on the price-quality heuristic (e.g. Rao & Monroe, 1989). Similar to the salience of opportunity costs in the last study, it does not appear plausible that the informative value of the reference price (i.e. higher reference prices suggesting a higher product quality) should depend on the payment mode.

In a different vein, the paradigm employed in this experiment can give further support for the findings of Raghubir and Srivastava (2009). More specifically, these authors found evidence for a "denomination effect" such that decision makers prefer to make payments with larger bills. In line with this phenomenon, participants in the cash condition selected the payment option highest denominations in 75% of the purchases.

Primarily however, the study aimed at an extension of previous research exploring the evaluative and behavioral consequences of different payment modes. However, while a general aversion towards highly transparent transactions has been documented before (Raghubir & Srivastava, 2008; Thomas et al., 2011), the current results point towards boundary conditions for this phenomenon. Specifically, evaluations were not unconditionally more negative if transactions were made in cash. Instead, the hypothesis that more transparent cash payments increase the weight given to Monetary Arithmetic is tentatively supported by the marginally significant interaction. To a certain degree, these results may even stimulate speculations about a certain "joy of paying". That is, if the sale price is considerably lower than the reference price, highlighting payments might lead to more positive evaluations of a transaction.

However, the results from the previous studies may

also have supported speculations in the opposite direction. In particular, paying in cash could highlight scarcity because the outflow of limited resources might be more salient compared to payments with credit cards (see Spiller, 2011). As a consequence, taking opportunity costs into account should be more likely if transactions are more transparent in this regard. Nonetheless, the previous study showed that comparing categorical evaluations of a transaction's benefits with their opportunity costs in fact decreased the impact of Monetary Arithmetic. However, emphasizing overt costs in monetary payment systems had the opposite effect of highlighted opportunity costs. Therefore, opportunity costs and consumptive benefits might generally play a more subordinate role in purchase decisions involving money. Maybe, the notion of Transaction Utility introduced by Thaler (1983) is a more appropriate interpretation of evaluations based on comparisons in the monetary dimension. Specifically, highlighting payments seems to replace comparisons of categorical evaluations by comparisons indicating a "good deal".

Unfortunately, if evaluations based on money render fundamental principles of economic decision making inoperative, this has wide ranging consequences for the interpretation of behavior investigated in settings abstracting from actual goods. Therefore, the following studies explore the evaluative principles underlying decisions in one of the most prominent paradigms in experimental economics: the Ultimatum Game.

4.5 Difficulty of Comparisons in the Ultimatum Game

The Ultimatum Game (Güth et al., 1982) ostensibly questions traditional Expected Utility Theory (e.g. von Neumann & Morgenstern, 1944/1955) and is often seen as the most prominent anomaly in behavioral economics (Thaler, 1988). The game is played between a proposer and a responder who negotiate how to split a given monetary "cake" C. Proposers offer a certain amount R to responders who may either accept or reject. If they accept, responders receive R and proposers receive $P = C - R$, i.e. the rest of the cake. Otherwise, no one receives anything. If responders reject non-zero offers, their choice is considered irrational, simply because receiving something is always better than receiving nothing. Nonetheless, numerous variations of the game have shown a positive acceptance gradient, which implies that acceptance depends on the *relative* size of the offer (Camerer, 2003a).

The "irrational" rejection of inequitable proposals has typically been considered to be a retaliatory response to an unfair offer (Fehr & Gächter, 2002). Although the typical Ultimatum Game is a one-shot exchange, it has been argued that such reciprocity is an important factor in the reinforcement of contracts and social norms (e.g. Bolton & Ockenfels, 2000; Falk & Fischbacher, 2006). In general, reciprocity accounts assume a matching of decisions between interaction partners. In particular, if one party cooperates the other party will also cooperate, but if the first party acts hostile, then the second party will retaliate (see also Axelrod & Hamilton, 1981). Applied to the Ultimatum Game, cooperation translates into equal divisions of the cake for the proposer and into accepting the division

by the responder. In turn, making unequal proposals is considered hostile while rejecting offers corresponds to retaliation. That is, the reciprocity explanation for the Ultimatum Game states that "fair" proposals are met with acceptance, while "unfair" proposals are met with rejection. Although the reciprocity approach describes the determinants of acceptance and rejection, it remains silent about the exact psychological process which underlie the tendency to reciprocate.

Instead, "irrational rejections" are sometimes considered impulsive and affect-based reactions which may be overridden by reflective processes (Camerer, 2003b; Hewig et al., 2011; Nowak, Page, & Sigmund, 2000; Sanfey et al., 2003; van't Wout, Kahn, Sanfey, & Aleman, 2006). In fact, patients with lesions in brain areas linked to emotion regulation have shown increased rejection rates (Koenigs & Tranel, 2007), as have participants who were deprived of sleep (Anderson & Dickinson, 2010) or were intoxicated by alcohol consumption (Morewedge, Krishnamurti, & Ariely, 2014). Similarly, participants whose cognitive capacity was reduced by lowering their serotonin levels (Crockett, Clark, Tabibnia, Lieberman, & Robbins, 2008) were more likely to reject such offers, as were males with high testosterone levels (Burnham, 2007). However, disrupting brain areas linked to executive function led to lower rejection rates (Knoch et al., 2006) which rather suggests that rejection may not be an impulsive, but a reflective response (Alos-Ferrer & Strack, 2014; Mussel, Göritz, & Hewig, 2013; Strack & Deutsch, 2004). Moreover, chimpanzees playing the Ultimatum Game accepted any offer (Jensen, Call, & Tomasello, 2007). Thus, they appear to act more rationally than humans. In sum, the evidence for a causal role of impulsive or emotional factors in the rejection of Ul-

timatum Game offers seems inconclusive. Therefore, it may be promising to investigate the appraisal process preceding affective reactions instead of focusing on the elicitation of affect itself.

In this research, we focus on the reflective component of decision making and assert that the acceptance of non-zero offers is more likely based on a categorical evaluation whereas rejections require a comparative judgment. That is, uCat is positive for all non-zero offers as they always contribute to consumptive goals. This proposition is entirely in line with traditional economic theorizing that the marginal utility of money is always positive (M. Friedman & Savage, 1948). In contrast, uCom decreases as the offers deviate from a standard of equal distribution (Camerer & Thaler, 1995; Güth, Huck, & Müller, 2001; Kahneman, Knetsch, & Thaler, 1986). In sum, whereas uCat prescribes the acceptance of any non-zero offer (acceptance gradient is zero), uCom commends the rejection of offers that are too far below the standard (acceptance gradient is positive). As a consequence, offers below the standard elicit a conflict between both evaluations. Moreover, the acceptance gradient should vary with the relative weights of both components.

In the following study, the impact of uCom was decreased by increasing the difficulty of the arithmetic operation required to determine how much a given offer deviates from a reference point. Therefore, we hypothesized that the acceptance gradient, which is interpreted as a manifestation of uCom, will decrease if the comparative judgment is more difficult. In a somewhat related vein, Handgraaf, Dijk, Wilke, and Vermunt (2004) have shown that if the payoffs of proposers and responders are different types of lottery tickets (similar to the P-bet and the \$-bet in Lichtenstein & Slovic,

1971) then the acceptance rate of substandard offers increases - which is equal to a decreased acceptance gradient. However, we opted for a more direct manipulation of difficulty which does not depend on the individuals' level of risk aversion. In addition, the approach underlying this study focused specifically on responders' reactions to systematically varied types of offers and did therefore not include actual proposers. However, in line with game theoretic reasoning, proposer behavior is largely determined by the expectations about how responders will react to each possible offer (for a more extended discussion see Weg & Zwick, 1994; Forsythe, Horowitz, Savin, & Sefton, 1994). Additionally, subjects expected responding to randomly selected proposers also sitting in the room.

Sample and design

Ninety-four participants (mean age $= 27.41$, $SD = 9.17$, 69% female) were invited for a laboratory study on decisions and behavior. They were informed that their compensation consisted of a fixed amount (3 Euros) plus their earnings from a negotiation task (1% of their total earnings; on average about 3.17 Euros).

Participants were tested in a 2 (ease of comparison: easy vs. hard; between subjects) $\times$ 5 (offer: 10% vs. 20% vs. 30% vs. 40% vs. 50%; within subjects) mixed design. Response time (RT) and acceptance rates were averaged across relative offers and served as dependent variables.

Procedure and materials

Upon their arrival, participants (four to eight in one room) were randomly assigned to the easy versus difficult conditions and seated in front of a computer in

a cubicle. Each computer was ostensibly connected to a network allowing the participants to play against each other. Participants were informed about the rules of the game, that they had been assigned the role of the responder and that they would be randomly paired with a different proposer for each round. Participants played a training round which secured that they understood the rules of the game. Unbeknownst to the participants, the offers were predetermined such that a randomized sequence of 150 offers (30 per level) was presented with a random time lag (1-5s) preceding each offer. In the easy condition, we replicated the structural elements of standard Ultimatum Game offers including pie diagrams, the proposed share (scaled in units of 100) and the total cake size of 1000 Euro cents per trial. In the difficult condition, offers were ostensibly made in different currencies that had been automatically converted into Euro cents, which resulted in odd cake sizes and offers. That is, the comparison with the standard as well as the equal split itself were harder to calculate. No pie diagrams were presented in this condition (see Appendix for details). Because the total amount offered over all trials had to be held constant between the ease of comparison conditions, the cake sizes in the hard-to-compare conditions varied (718ct, 872ct, 936ct, 1064ct, 1128ct, 1282ct). After playing 150 trials, the total earnings were displayed and paid out.

Results

The data was analyzed by estimating the β-parameters for two independent multi-level linear mixed-effect models:

$$\begin{aligned} log(RT) = (\beta_0 + \beta_1 H) \\ +O * (\beta_2 + \beta_3 H) \\ +O^2 * (\beta_4 + \beta_5 H) + Res. \end{aligned} \tag{4.3}$$

$$\begin{aligned} A = (\beta_0 + \beta_1 H) \\ +O * (\beta_2 + \beta_3 H) + Res. \end{aligned} \tag{4.4}$$

The dependent variables "response latency" and "acceptance rate" are denoted by RT and A, respectively. H is a dummy variable for the difficult-to-compare condition. O is a numeric predictor variable denoting the relative offer.

Reaction Times

Table 4.4 shows the maximum likelihood estimates for the parameters of the response latency model described by Equation 4.3.

Variable	i	β_i	*SD*	*df*	*p*-Value
Intercept	0	-0.9847	0.40711	184	0.017
H	1	0.8027	0.57575	92	0.167
O	2	0.0808	0.02850	184	0.005
O x H	3	-0.0424	0.04030	184	0.295
O^2	4	-0.0014	0.00047	184	0.004
O^2 x H	5	0.0008	0.00067	184	0.262

Table 4.4: Estimates for the response latency model described by Equation 4.3.

The significant parameters β_0, β_2 and β_4 show that response latencies are well described by a parabola across offer levels. Because β_4 was negative, the data pattern shows an inverted U-shaped curve such that extreme offers prompted fast reactions whereas decisions concerning the intermediate offers took longer (see Figure 4.9). Thus, response latencies were low for 10% offers, increased towards 30% offers and fell again for 50% offers. In line with our prediction, the delayed responses for intermediate offers suggest conflicting evaluations while there is less of a conflict for 50% and 10% offers.

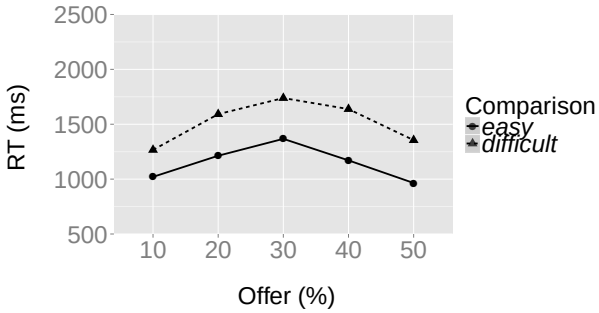


Figure 4.9: The figure shows the inverted U-shaped relationship between response latencies and relative offers indicating that intermediate offers required more time to respond. Also, latencies for difficult-to-compare offers were generally longer.

In addition, the positive parameter β_1 indicates a vertical upward shift of the parabola for difficult-to-compare offers but the main effect in the regression failed to reach significance ($p = 0.167$). However, collapsed over all offer levels, differences in response times were highly significant such that responses in the difficult condition took 32% longer than in the easy condition ($t(84.60) = 2.79$, $p = 0.006$).

Acceptance Rates

Figure 4.10 shows that acceptance rates are best described by a sigmoid curve, however, the acceptance rate model described by Equation 4.4 was estimated using only the intermediate offers (20%, 30% and 40%).

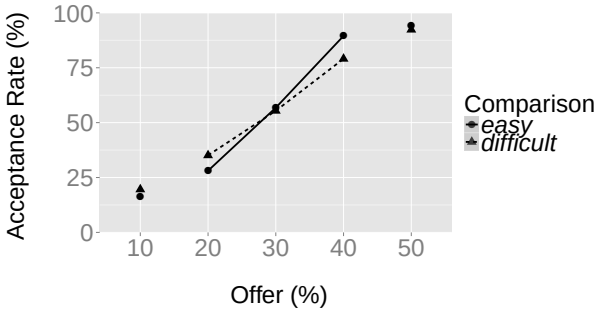


Figure 4.10: The figure shows the sigmoid relationship between acceptance rates and relative offers and the linear approximation for intermediate offers. Most important, the slope decreases when comparisons were difficult.

We limited our analysis to these offer levels to remove floor and ceiling effects and because a simple linear model is a feasible approximation in that range. We are aware that more complex statistical modeling would allow a more exact description of the curve. However, the necessary inclusion of the extreme offers would make the interpretation of the results less intuitive. Also, we did not intend to identify the formal relationship between offers and acceptance rates. Instead, we attempted to show that the Ultimatum Game anomaly, that is, the positive acceptance gradient captured by the slope of the linear approximation, is affected by our experimental manipulations.

Table 4.5 shows the maximum likelihood estimates for the linear approximation. Obviously, the significant but negative estimate for β_0 cannot be interpreted as

acceptance rates can never be negative. The same is true for the intercept of the difficult-to-compare curve which results from adding the significant estimate for β_1 to β_0. In contrast, the significant and positive estimate for β_2 unambiguously shows that the acceptance gradient is in fact positive: as the relative offer was increased by 1%, the acceptance rate on average increased by 3.07%. Most important, the significant and negative estimate for β_3 denotes the decrease of the acceptance gradient for difficult-to-compare offers. In fact, increasing these relative offers by 1% only led to an increase of 2.20% in acceptance rates. Thus, the acceptance gradient was 28% weaker in the difficult condition.

Discussion

The results from this experiment shed light on the cognitive dynamics of the Ultimatum Game. In general, the predictions of the Dual Utility framework are confirmed by the findings. Firstly, the increased reaction times for intermediate offers point to the presence of a conflict between the evaluative implications of com-

Table 4.5: Parameter estimates for the acceptance rate model described by Equation 4.4 from Study 5.

Variable	i	β_i	*SD*	*df*	*p*-Value
Intercept	0	-34.00	8.14	186	< 0.001
H	1	24.56	11.51	92	0.036
O	2	3.07	0.24	186	< 0.001
O x H	3	-0.87	0.34	186	0.010

parative versus categorical utility. Therefore, we propose that before responders accept an offer they must decide whether uCat outweighs uCom. In particular, while uCat and uCom do not have opposing implications for 50% offers, the uCom driven propensity to reject 10% offers presumably dominated any concerns about uCat. Correspondingly, extreme offers exhibited both, a low degree of response ambiguity (i.e. 10% offers were rejected most of the time whereas 50% offers were accepted almost every time) and fast reaction times. As a consequence, the intensity of the conflict appears rather low for extreme offers but increased for intermediate offers. In sum, these results encourage future investigations of reaction times in the Ultimatum Game to yield a deeper insight into the underlying cognitive processes.

Secondly, the acceptance gradient was shown to hinge on and the ease of generating proportional judgments. Specifically, difficult-to-compare offers led to a smaller acceptance gradient. Even though the relative shares received, i.e. the outcomes for decision maker were identical in the easy- and difficult-to-compare conditions, the judgments leading to this conclusion could probably not be made with the same precision. Therefore, we suggest that this variation in the game's setting shifted utility focus away from uCom which decreased the sensitivity of behavior to the relative size of the offer. In general, we propose that these results are best explained in terms of an interplay between the described evaluative processes.

To be sure, monetary payoffs are quite common incentives in experimental economics. However, they make for poor stimuli regarding a categorical evaluation. In particular, except largely irrelevant characteristics like the currency, the only attribute of money is

its amount. Therefore, more elaborate evaluations of the money in terms of uCat may not be very likely. For example, thoughts about its consumptive potential for desired purchases may not readily come to mind. Simultaneously, the opportunity costs of foregoing consumption might not have been considered as well. As a consequence, the uCom driven acceptance gradient would have been inflated. In the next experiment, we attempt to shift utility focus towards uCat by highlighting the consumptive consequences.

4.6 Inducing Gossen's Accounting in the Ultimatum Game

The last experiment was designed to replicate the previous findings and to direct participants' attention toward the consumptive consequences of their decisions.[12] This was achieved by converting subjects' earnings into vouchers of a store they had previously selected. Importantly, this alteration should not increase the desirability of the reward which should actually decrease because of the limited fungibility. Instead, we hypothesized that the conversion would change the way subjects thought about the reward. In detail, converting earnings into gift vouchers was intended to increase the salience of the opportunity costs of rejection. That is, we hypothesized that, in addition to the contextual reference point (i.e. the equal distribution), the "rational" comparison between opportunity costs

[12]Interestingly, the terminology which evolved in relation to the Ultimatum Game suggests that the payoffs in the game actually do possess some experiential qualities. In particular, Güth (1995) introduces the game in terms of the experientially very rich activity of eating: "We refer to c as to the 'cake' which X and Y can 'eat.'" (p. 330).

and marginal utility would also influence the decision. In particular, rejecting a positive offer inflicts opportunity costs equal to the foregone consumptive pleasures while the marginal utility of 0 Euro is zero. In contrast, accepting any positive offer yields marginal utility corresponding to those consumptive pleasures without any opportunity costs.

Sample and design

One-hundred-twenty-six participants (mean age $= 24.56$, $SD = 6.79$, 71% female) were invited for a laboratory study on decisions and behavior and were told at the outset that they would receive 2.50 Euros plus 1% of their earnings from a negotiation task (on average about 5.51 Euros).

Subjects were tested in a 2 (ease of comparison: easy vs. hard; within subjects) $\times$ 5 (relative offer: 10%, 20%, 30%, 40%, 50%; within subjects) $\times$ 3 (offer framing: control, coupon, coupon & sum; between subjects) mixed design. Again, reaction times and acceptance rates, averaged across within-levels, served as dependent variables.

Procedure and materials

Before playing the Ultimatum Game, participants completed a short task to systematically highlight the consumptive consequences of money. In the experimental conditions they were shown four logos of well-known stores with local branches (see Appendix for details) and were asked to think for two minutes about possible purchases. Then, participants had to select the store from which they would like to obtain a gift voucher into which their earnings would be converted. In the control condition, the earnings were offered in cash

and participants were presented city signs (see Appendix) and asked to estimate the size of the cities. Payment across conditions was identical except that in the experimental conditions participants expected a gift voucher instead of cash. The remaining procedure was the same as in the first experiment except that 240 rounds were played, no pie diagrams were used and the selected store logo was displayed with every single offer in the experimental conditions. In addition, participants in the "coupon & sum" condition saw the amount of money they had already accumulated for their gift certificate (see Appendix for details).

At the very end of the experiment, all participants received an envelope containing their earnings from the Ultimatum Game plus the show-up fee. Participants in the experimental conditions who were expecting a gift voucher received an envelope with the logo of the selected store and the word "Gutschein" (gift voucher) printed on. Participants in the control condition received their money in a blank envelope.

Results

Again, we estimated the β-parameters for two independent multi-level linear mixed-effect models:

$$\begin{aligned} log(RT) = (\beta_0 + \beta_1 C + \beta_2 CS + \beta_3 H) \\ +O * (\beta_4 + \beta_5 C + \beta_6 CS + \beta_7 H) \\ +O^2 * (\beta_8 + \beta_9 C + \beta_{10} CS + \beta_{11} H) + Res. \end{aligned} \tag{4.5}$$

$$\begin{aligned} A = (\beta_0 + \beta_1 C + \beta_2 CS + \beta_3 H) \\ +O * (\beta_4 + \beta_5 C + \beta_6 CS + \beta_7 H) + Res. \end{aligned} \tag{4.6}$$

In addition to the variables from Equation 4.3 and 4.4, C and CS are the new dummy variables for the coupon and coupon & sum conditions, respectively.

Reaction Times

Table 4.6 shows the maximum likelihood estimates for the quadratic model formalized by Equation 4.5.

Variable	i	β_i	*SD*	*df*	*p*-Value
Intercept	0	6.6588	0.08070	627	< 0.001
C	1	0.1941	0.10846	123	0.076
CS	2	0.2986	0.10718	123	0.006
H	3	-0.0207	0.05294	627	0.70
O	4	0.0196	0.00468	498	< 0.001
O x C	5	-0.0133	0.00602	498	0.028
O x CS	6	-0.0151	0.00594	498	0.011
O x H	7	0.0139	0.00403	627	0.001
O^2	8	-0.0003	0.00008	498	< 0.001
O^2 x C	9	0.0002	0.00010	498	0.031
O^2 x CS	10	0.0002	0.00010	498	0.056
O^2 x H	11	-0.0002	0.00007	627	0.004

Table 4.6: Parameter estimates for the response latency model described by Equation 4.5 from Study 6.

Again, the significant parameters β_0, β_4 and β_8 support the notion of a quadratic relationship between relative offers and response latencies. Furthermore, the significant and negative estimate for β_8 implies an inverted U-shaped pattern similar to the first study

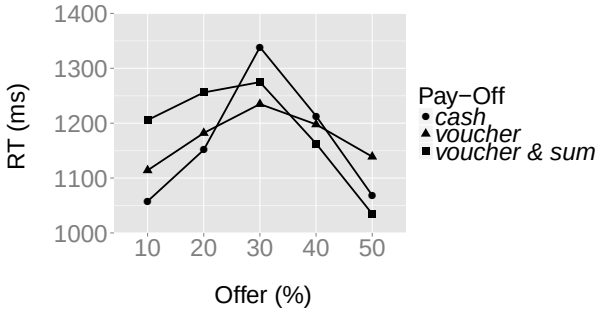


Figure 4.11: The figure shows response latencies across offers. Importantly, highlighting consumptive consequences attenuates the differences between extreme and intermediate offers.

(see Figure 4.11). The parameter estimates for β_7 and β_{11} were significant which means that the parabola is moved upwards for difficult-to-compare offers. Aggregated over all offer levels, responses in the difficult condition took 25% longer than in the easy condition ($t(125) = 11.16,\ p < 0.001$).

In addition, the estimates for β_1, β_5 and β_9 as well as for β_2, β_6 and β_{10} indicate that the parabola is widened in the coupon and coupon & sum conditions which did not differ from each other. That is, compared to the control condition, where reactions to extreme offers were faster than to intermediate offers, the wider spread of the parabola was caused by slower reactions to extreme offers in the experimental conditions.

Acceptance Rates

Table 4.7 shows the maximum likelihood estimates for the linear model described by Equation 4.6. Please note that we again limited our analysis to the intermediate offers and will not further discuss the intercept parameters. The parameter estimate for β_4 was positive which indicates a positive acceptance gradient. The negative

Table 4.7: Parameter estimates for the acceptance rate model described by Equation 4.6.

Variable	i	β_i	*SD*	*df*	*p*-Value
Intercept	0	-47.41	8.16	376	< 0.001
C	1	32.71	11.37	123	0.005
CS	2	49.85	11.24	123	< 0.001
H	3	2.73	3.27	376	0.40
O	4	3.55	0.24	249	< 0.001
O x C	5	-1.04	0.33	249	0.002
O x CS	6	-1.01	0.33	249	0.002
O x H	7	-0.29	0.11	376	0.006

estimate for β_7 is consistent with the previous study and implies a decrease of the acceptance gradient in the difficult-to-compare condition.

Figure 4.12 illustrates the acceptance rates for the different framings of the offer. As predicted, and parallel to increasing the difficulty of comparisons, the analysis of the linear approximation for intermediate offers showed a much weaker acceptance gradient when consumptive consequences were highlighted. Specifically, the estimates for β_5 and β_6 were negative suggesting that the acceptance gradient was diminished by highlighting consumptive consequences (by 29% in the coupon condition and 28% in the coupon & sum condition). In the control condition the acceptance rates on average increased 3.55% for a 1% increase in relative offers. In contrast, 1% increase in relative offers generated a 2.51% increase in acceptance rates in the coupon condition and a 2.54% increase in the coupon

& sum condition.

Averaged across all offer levels, 56.2% were accepted in the control condition and 58.5% in the coupon condition ($t(75.32) = 0.47$, $p = 0.638$). In contrast, 70.1% of the offers were accepted in the coupon & sum condition which is higher than in the coupon condition ($t(82.53) = 3.67$, $p < 0.001$) and in the control condition ($t(72.37) = 2.64$, $p = 0.001$). Therefore, general acceptance rates were increased only if the total of all earnings was salient.

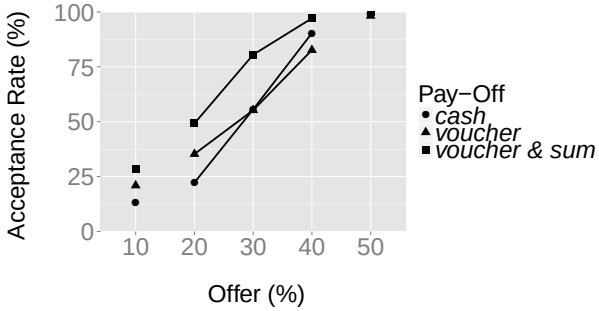


Figure 4.12: The figure depicts acceptance rates across offers for each offer framing. The linear approximation for intermediate offers indicates a decrease in slopes when consumptive consequences are highlighted.

Discussion

To begin with, the results from the previous study have all been replicated. In detail, the inverted U-shaped offer-reaction time relationship suggesting a conflict between uCat and uCom was also supported by the present findings. Moreover, difficult-to-compare offers again took longer and yielded a smaller acceptance gradient.

As an extension of the previous findings, increasing the salience of consumptive consequences attenuated

the inverted U-shape of the reaction times. In fact, both experimental conditions show almost no difference between extreme and intermediate offers.[13] Unfortunately, the current results do not allow more precise conclusions about the differences in reaction times for each offer levels because contrasts were not significant. To be sure, shifting utility focus towards uCat should influence the shape of the reaction time curve. Specifically, extreme offers far below the standard should imply increased conflict and reaction time if the weight of uCat is increased whereas extreme offers close to the standard should not. That is, in the first case, the intensity of the conflict should increase because uCat is reinforced against a strong uCom component. In contrast, there is hardly any conflict in the latter case because uCom does not dictate a rejection and is therefore compatible with the strengthened uCat component. Nonetheless, the altered reaction time curve may be tentatively interpreted in favor of the evaluative conflict hypothesis.

More importantly perhaps, the acceptance gradient was decreased if the salience of consumptive possibilities was increased. Therefore, this alteration of the game setting had the same effect as increasing the difficulty of comparative computations. For once, these results can be conceptualized as shifts in utility focus away from uCom. In more detail, acknowledging the consumptive potential of money allows a multifaceted categorical evaluation of the offer. That is, instead of evaluating a naked number, which might almost inevitably trigger arithmetic operations, the of-

[13] Also, constructing the reaction time curves from the parameter estimates shows that the respective parameters for the quadratic ($\beta_8 + \beta_9$ for the coupon condition and $\beta_8 + \beta_{10}$ for the coupon & sum condition) and linear term ($\beta_4 + \beta_5$ for the coupon condition and $\beta_4 + \beta_6$ for the coupon & sum condition) almost cancel out.

fered money may serve solely as a dummy for some good which is subjected to a categorical evaluation. Also, these judgments about uCat could play a role as the opportunity costs of rejection framed as the consumptive pleasures that would have to be foregone. All the same, if the Monetary Arithmetic is thereby crowed out by a comparison in terms of uCat, then the acceptance gradient should also decrease. Therefore, these results are entirely in line with the proposed Dual Utility Model.

In sum, the present findings suggest that a positive acceptance gradient is less likely to be driven by affective impulses than by reflective judgments of uCom although emotional feelings may be elicited by the result of the comparison (see Sanfey et al., 2003). Regarding a causal role of affective forces, however, the intensity of affect should increase monotonically the more an offer deviates from the standard, but reaction times indicate a non-monotonic relationship. Therefore, this pattern cannot be explained by an increase in affect alone, but only by the parallel operation of two action tendencies elicited by two different evaluations. To be sure, emotions play an important role in many economic decisions, but this is true for both impulsive and reflective choices (Strack & Deutsch, 2004; Loewenstein & Lerner, 2003; Pham, 2007). Therefore, to predict players' reactions in the Ultimatum Game, it seems more useful to identify the interplay of different evaluative processes than the concomitant affective experience.

At the same time, the present notion is in line with the reciprocity account in that substandard offers are seen to be unfair (Fehr & Schmidt, 1999). But going beyond fairness and the Ultimatum Game, uCom also applies to outcomes above a reference point. For example, a "good deal" or "saving" through a discount

is also the result of comparison (Thaler, 1985/2008) and may also be accompanied by a pleasant experience (Schindler, 1989, 1998). Also, the results from the current study cannot be explained by the reciprocity account as it does not seem reasonable to assume that thinking about money in terms of consumption should attenuate the importance of fairness. More generally, we propose that affective component of "unfair" offers not only refers to anger about the proposers behavior (e.g. van't Wout et al., 2006), but that these feelings also reflect an evaluation of the responder that is linked to self-esteem. That is, in addition to the punishment and education of the proposer, our approach emphasizes the role of responders' self respect in the rejection of "bad deals". In a related vein, there is evidence (Gu, Bohns, & Leonardelli, 2013) that inducing prevention focus increases the acceptance gradient, i.e. relatively low offers are rejected more frequently in prevention focus. In line with our interpretation, persons who tend to prevent negative outcomes in general also prevent making bad deals in the Ultimatum Game.

Chapter 5

Conclusion

"Utility" is perhaps the most central concept in modern economic theorizing. However, the unified notion based on observable behavior (e.g. von Neumann & Morgenstern, 1944/1955; Samuelson, 1938) gave rise to numerous "anomalies" demonstrating an inconsistency between experimental findings and economic theory (Thaler, 1994). Simultaneously, the economic situation created by worldwide financial crisis largely resisted any attempts to be influenced or explained along the lines of traditional economic theories (see e.g. Shiller, 2015). At the center of both challenges to economic theorizing is the concept of utility. In this program of research, we focused on the psychological processes by which judgments of utility are generated. Specifically, we challenged the standard assumption of a homogeneous concept and proposed that two evaluative components must be separated. These components are psychologically distinct and jointly establish a dual notion of utility. In detail, utility is assumed to have a categorical and a comparative constituent.

Judgments concerning categorical utility (uCat) infer the category membership of a good based on its at-

tributes. Subsequently, semantic knowledge about the category (e.g. attitudes or contributions to the pursuit of a goal) can be applied to the object. Alternatively, categorical evaluations may resort to affective information to generate a judgment (e.g. Schwarz & Clore, 1983). In contrast, comparative evaluations (uCom) rely on the interpretation of the distance between a target and standard on a specific dimension of comparison. Even though economists generally consider only utility itself as a relevant dimension of comparison, comparing alternatives on different dimensions may not only diverge in their evaluative implications but also in the underlying psychological processes. Together, these components make a Dual Utility Model which provides a parsimonious explanation for a wide range of economic phenomena. The central predictions of this framework were tested in a series of six experiments.

The first study showed that anomalies together with corresponding explanations often arise because the psychological mechanisms of evaluation are not adequately taken into account. In particular, very specific social norms have been cited as the cause of anomalous behavior (Heyman & Ariely, 2004) instead of considering the lack of commensurable standards of comparison. However, after standards of comparison were made available, the anomaly vanished almost entirely. Therefore, taking a look at the evaluative processes promises more parsimonious explanations for economic behavior than resorting to rather ad hoc interpretations.

In the second study, we explored how evaluative comparisons affect behavior in the indigenous domain of the utility model. Specifically, we analyzed changes in resource allocation caused by a price shift. Accord-

ing to the traditional notion of utility, prices are solely a constituent of opportunity costs, i.e. uCat. At the same time however, a price might be compared with a reference price which would manifest in an additional uCom component. In line with this reasoning, the results of the experiment showed an amplified price effect on demand in a second period. Apparently, the price from the first period served as a standard providing a new dimension of comparison. As a consequence, this led to "overspending" if the reference price from the previous round was higher, and to "underspending" if it was lower. Thus, we proposed that this increased shift in demand may be explained by an additional uCom component hinging on the a comparison in the monetary dimension.

The third study addressed the interaction between two economically important dimensions of comparison. In detail, the results showed that comparisons in the monetary domain had a smaller impact on the evaluation of a transaction if opportunity costs were directly cued as an additional standard of comparison. Thus, different uCom components, which are each based on a specific dimension of comparison, may partly replace each other in the evaluation of transactions.

The fourth study investigated how increasing the salience of monetary costs affects the evaluation of transactions in terms of price comparisons. The proposed Dual Utility Model allowed two opposing predictions regarding those more transparent constructions of mental accounts (Thaler, 1985/2008). Firstly, highlighting costs might have directed decision makers' attention toward the scarcity of resources what might in turn have served as a reminder to consider opportunity costs (see Section 2.1). Therefore, based on the findings in the third study, uCat would be expected to partly

replace money as a dimension of comparison. Secondly, increased transparency concerning costs might also have obscured thoughts about the benefits of the transaction. To a considerable extent, transactions exhibiting this kind of transparency may therefore be evaluated based on the "merits of the deal" (Thaler, 1985/2008, p. 19). Thus, evaluations would be dominated by the monetary dimension. Interestingly, the results of the experiment supported this second prediction. That is, Monetary Arithmetic had a larger impact on the transaction evaluations if payments were more salient. To some degree, these findings suggest that Monetary Arithmetic may be decoupled from the consumptive benefits generated or rejected in a transaction and mainly refer to the terms of trade.

The utilitarian dynamics in the Ultimatum Game (Güth et al., 1982) were the focus of the fifth study. Specifically, we tested the hypothesis that the mere difficulty of the comparative operation may shift utility focus away from uCom. In the Ultimatum Game, offers tend to be rejected by the responding player as the deviation from an equal split increases. That is, if a proposer offers 50% of the sum total, this offer is most likely accepted by the responder. However, if the offer amounts to only 10%, it is most likely rejected. From the perspective of our Dual Utility Model, the sensitivity of the responders' behavior to the *relative* size of the offer is solely a manifestation of negative uCom. After all, every amount of money earned without any effort (as it is always the case in this game) would be assigned a positive uCat. In line with the prediction of the proposed model, the results of this study indicated that the acceptance gradient (a measure of the previously mentioned sensitivity) decreased if the computations of the offers' *relative* size was made more difficult.

In addition, the presence of conflicting evaluative implications from uCat and uCom was supported by an analysis of reaction times.

The last study explored another aspect of the reasoning developed in the previous experiment. This time, we highlighted the consumptive potential inherent in any amount of money. In line with the Dual Utility Model, bringing desired purchases to mind apparently triggered a more categorical evaluation of the offers. In turn, the impact of comparisons in the monetary dimension was reduced which manifested in a decreased acceptance gradient. Therefore, enriching a monetary payoff with the anticipation of actual consumptive experiences decreased the effect of arithmetic operations concerning quantities of money. As a mere number, money is one-dimensional and comparisons are restricted to monetary reference points. Conversely, if money is processed as a wildcard for consumption, categorical evaluations of the goods behind the money may open up many more dimensions of comparison. For example, instead of rejecting a 10% offer due to a comparison with a 50% offer, it might be accepted if it is processed as a cup of coffee which would have to be foregone if the offer was rejected. Therefore, while Monetary Arithmetic probably dominates decisions in standard economic games, Gossen's Accounting, and therefore decisions corresponding to the traditional models, may emerge if money is actually construed as a mean for consumption.

The findings from the last two studies particularly illustrate that the one-dimensionality of money is to some degree responsible for the inconsistencies between the standard notion of utility and behavior in experimental settings. That is, a major share of experimental economists investigated decisions incentivized

with monetary payoffs thereby excluding a considerable part of the evaluative processes preceding behavior. Specifically, the psychological nuances regarding different dimensions of comparisons have been ignored entirely. As a consequence, experimental situations solely featuring monetary incentives inflate the role of Monetary Arithmetic for evaluation and behavior. The findings of the current studies suggest that utility transcends money which implies that the simplified settings in the laboratory may not be an adequate basis to generalize the findings to decisions outside the laboratory. That is, if the environmental validity of the findings should not be sacrificed to simplification, then consequences must to be drawn for experimental practice. To be sure, M. Friedman and Savage (1948) had the construction of theoretical models in mind when they promised that no generality would be lost if utility would be equated with money. Nonetheless, monetary payoffs became the Gold Standard in economic experiments and up until now few attempts have been made to correct for this ill-advised simplification.

On a different level of analysis, the cumulative evidence from all experiments has far reaching implications for the interpretation of economic behavior. Specifically, it is impossible to infer the exact nature and content of an evaluative judgment from observations of the subsequent behavior. Nevertheless, following modern theories of utility (e.g. Hicks & Allen, 1934a; Samuelson, 1938), the measure of subjective evaluations is to be found in the relation between behavior and the exchange conditions in the market. That is, for the existence of utility it is only necessary that behavioral reactions to changes in prices and income meet specific consistency requirements (see section 2.2 and 2.3). However, even if choices satisfy these criteria

(see also Varian, 2006), and the existence of a utility function is therefore guaranteed, it is still not possible to make any definite claims about the interpretation of this function. The evidence presented here suggests at least two possible interpretations concerning the pivotal observation of a specific transaction in the current market situation. First, it is possible that all costs and benefits have been subjected to a categorical evaluation creating a dimension for a comparative evaluation of the transaction. Therefore, the observation might be interpreted such that the uCat difference between benefits and costs is largest for the observed transaction. Certainly, this would be an interpretation favored by a majority of economists. At the same time, we have presented abundant evidence that uCat is not the only dimension for comparative judgments. Therefore, the above observation may also prompt a second interpretation such that the decision maker deems the observed transaction a good or acceptable deal. Importantly, such an evaluation does not rely on uCat as a dimension but may solely stem from another uCom component based on a comparison of the current price with a reference price.[1] Therefore, no observation of a transaction can clarify the dimension used to compare costs and benefits nor is it possible to claim the occurrence of a comparison altogether. In a related vein, Kahneman, Wakker, and Sarin (1997) also proposed that the proto-behaviorist approach where utility is inferred from decisions ("decision utility") does not account for the actual "experienced utility" (see also Kah-

[1] It shall be noted here that even in the absence of a formal investigation, the intuition seems justifiable that any uCom component that assigns a positive value to transactions in which the reference price is higher than the actual price suffices to secure both, consistency of behavior (see Ariely et al., 2003) and the law of demand (e.g. G. S. Becker, 1962).

neman & Thaler, 2006). Nonetheless, economists often prematurely choose the interpretation of comparisons in terms of experiences or, more generally, uCat which demonstrates a certain "Utilitarian Fallacy".

Common manifestations of the Utilitarian Fallacy are distorted expectations concerning certain measurements. That is, data is often classified as anomalous because researchers erroneously expect to find a match between measures of behavior and those capturing the happiness or satisfaction with the outcome of the decision even though the underlying evaluative processes may be very different. For example, Tversky and Griffin (1991) have shown that participants predict to be happier in a job with a lower absolute income but where they earn more than their peers. In contrast, if the same participants are asked to predict which job they would actually choose, they chose the one with the higher absolute income but where they would earn less than their peers. In the light of the Utilitarian Fallacy, these results are hardly surprising because the standard of comparison concerning the decision probably is the own wage in the other job. In contrast, the judgment about happiness in one position does not rely on a comparison with the other. Instead, participants might have engaged in some kind of affective forecasting (Gilbert et al., 2002) about working in an environment where they are outearned by all the co-workers and subsequently derived a corresponding categorical evaluation (for another example see Hsee, 1999). Apparently, the misconceptions surrounding the Utilitarian Fallacy can be avoided by partly reversing the Paretian Turn (see Bruni & Sugden, 2007) and reintroducing the psychological content of the evaluations preceding observable decisions. Based on choices alone, utility is a merely formal construct without any inherent

meaning. At the same time, any ex post introduction of meaning runs the risk of falling prey to the Utilitarian Fallacy.

Furthermore, the established approach towards the Ultimatum Game also shows signs of the fallacy. Specifically, previous investigations to a large degree imply that rejections inflict opportunity costs of foregoing consumption and are therefore irrational because this behavior yields no obvious benefits. Starting with this irrationality, rejections of non-zero offers are usually interpreted as an anomaly caused by affective forces (Sanfey et al., 2003; Knoch et al., 2006). However, following the interpretation of our Dual Utility Model, rejections are caused by a comparative evaluation of the "deal" that is reflective in nature (cf. Strack & Deutsch, 2004). Even though the consequences of uCom appear to be the opposite of greed in the setting of the game (after all, subjects renounce additional income due to comparative operations), the explanations for seemingly greedy bankers often also focus on emotional and impulsive factors instead of recognizing the idiosyncrasies of reflective evaluations. That is, a CEO switching positions because a new job offer promises ten million instead of nine probably based the decision on a simple comparative evaluation. However, an observer of this behavior may not compare ten million to nine but to the own income of 60.000. From this perspective, the new position does not appear to be very superior which opens the stage for other attempts to explain the decision. For example, the CEO might be considered completely immoderate and possessed by impulses of greed.

To be sure, a certain class of anomalies is correctly attributed to impulsive factors. However, the studies summarized in this thesis suggest that a large part

of anomalous or even "irrational" behaviors may be explained more parsimoniously by exploring the psychological processes underlying evaluative judgments. Therefore, the proposed utilitarian duality may be a good starting point for a more general conceptual framework which integrates the assumptions of both the "rational" model and its anomalies. As a consequence, anomalies may become less important as mere examples for the shortcomings of economic models (Levine, 2012). Instead, they may become testing grounds for any theorizing that takes a closer look at the laws of utilitarian assessment.

Without any doubt, more research is necessary. In general, future investigations might address two different types of questions. First, the current studies to some degree focused on the psychological mechanisms underlying uCom. Therefore, the formation of categorical evaluations and their role in paradigms indicating anomalous behavior should be investigated. For example, it can be hypothesized that the Asymmetric Dominance Effect (Huber et al., 1982), which is most likely driven by a facilitation of comparative evaluations (see Section 3.2), may be overcome or attenuated if characteristics of the stimulus shift peoples' utility focus towards uCat. Specifically, we assume that more vivid stimuli elicit stronger experiential reactions (Strack et al., 1985) which may be used to generate categorical judgments (Schwarz & Clore, 1983). Consequently, vivid stimuli should attenuate the effect of asymmetric dominance. In addition, the results from the third study showed that the neglection of opportunity costs (Frederick et al., 2009) may be overcome by cuing alternative uses of resources. Moreover, we hypothesized that mental simulations (e.g. Adaval & Wyer, 1998; Elder & Krishna, 2012) may be used to generate in-

formation for categorical evaluations. Therefore, the crowding out of comparisons in the monetary dimension by those in terms of uCat (i.e. comparisons with opportunity costs) may be increased if the opportunity cost reminders are more vivid and trigger more elaborate mental simulations. For example, the text-based cues used in the third study could be replaced by pictorial cues.

A second direction of research might extend the current framework to include motivational factors influencing utility focus. In particular, we propose that uCom is partially fueled by internal norms of perceived self-efficacy (Bandura, 1991). That is, a price comparison indicating a good deal may nurture the conviction of having "outsmarted" the market which may be reflected in increased self-esteem. In turn, negative feelings about the self may be elicited after learning that the neighbor struck a better bargain at the gas station. Therefore, if self-esteem is threatened, utility focus might shift toward uCom in an effort to compensate. Conversely, realizing a comparative advantage in a transaction might boost self-esteem. These questions could also be addressed in the asymmetric dominance paradigm (see Huber et al., 1982). Specifically, threatening self-esteem (e.g. via false performance feedback) should increase the Asymmetric Dominance Effect. By the same logic, participants choosing asymmetrically dominant options should show an increase in self-esteem compared to those who chose the same option without a decoy being present.

Finally, the present psychological reformulation of utility needs to be reintegrated into economic modeling. The fields of application are as manifold as the roles played by the traditional utility paradigm in these models. A particularly interesting and provoking ap-

plication may be inspired by the classical findings of (Easterlin, 1974). The empirical fact that economic growth is not reflected in the well-being of the population beyond a certain GDP per capita is strongly at odds with the traditional theory of utility. After all, GDP measures the number of transactions in an economy weighted by their monetary values. Because any transaction should only take place if it increases the utility of at least one party (Pareto Optimality (see Lockwood, 2008)) and utility is traditionally equated with happiness (see Read, 2007; Yang, 2001), more transactions should be synonymous with more happiness. In contrast, recognizing the dual nature of utility shows Easterlin's paradox in a new light. Some transactions are solely made on the basis of uCom but their positive evaluation may fade quickly as the initial standards of comparison are forgotten. As a consequence, the decision maker has contributed to economic growth but may not be much happier after purchasing the discounted toaster.

References

Adaval, R., & Wyer, R. S. (1998). The role of narratives in consumer information processing. *Journal of Consumer Psychology*, *7*(3), 207–245.

Afriat, S. N. (1967). The construction of utility functions from expenditure data. *International economic review*, *8*(1), 67–77.

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179–211.

Akerlof, G. A., & Shiller, R. J. (2010). *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism*. Princeton University Press.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society*, 503–546.

Allen, R. G. (1932). The foundations of a mathematical theory of exchange. *Economica*(36), 197–226.

Alos-Ferrer, C., & Strack, F. (2014). From dual processes to multiple selves: Implications for economic behavior. *Journal of Economic Psychology*, *41*, 1–11.

Amir, O., Ariely, D., & Carmon, Z. (2008). The dissociation between monetary assessment and predicted utility. *Marketing Science*, *27*(6), 1055–1064.

Anderson, C., & Dickinson, D. L. (2010). Bargaining and trust: the effects of 36-h total sleep deprivation on socially interactive decisions. *Journal of sleep research*, *19*(1-Part-I), 54–63.

Antonelli, G. (1886). *Sulla teoria matematica dell economia politica*. Edizioni Fochetto, Pisa.

Ariely, D., & Jones, S. (2008). *Predictably irrational*. HarperCollins New York.

Ariely, D., & Loewenstein, G. (2006). The heat of the moment: The effect of sexual arousal on sexual decision making. *Journal of Behavioral Decision Making*, *19*(2), 87–98.

Ariely, D., Loewenstein, G., & Prelec, D. (2003). "coherent arbitrariness": Stable demand curves without stable preferences. *Quarterly Journal of Economics*, *118*(1), 73–106.

Ariely, D., Loewenstein, G., & Prelec, D. (2006). Tom sawyer and the construction of value. *Journal of Economic Behavior & Organization*, *60*(1), 1–10.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396.

Bagwell, L. S., & Bernheim, B. D. (1996). Veblen effects in a theory of conspicuous consumption. *The American Economic Review*, 349–373.

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational behavior and human decision processes*, *50*(2), 248–287.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, *51*(6), 1173.

Baumeister, R. F. (2002). Yielding to temptation: Self-control failure, impulsive purchasing, and consumer behavior. *Journal of Consumer Research*, *28*(4), 670–676.

Becker, G. S. (1962). Irrational behavior and economic theory. *The Journal of Political Economy*, 1–13.

Becker, G. S. (1965). A theory of the allocation of time. *The economic journal*, 493–517.

Becker, G. S. (1973). A theory of marriage: Part i. *The Journal of Political Economy*, 813–846.

Becker, G. S. (1974). A theory of marriage: Part ii. *The Journal of Political Economy*, 11–26.

Becker, G. S., & Murphy, K. M. (1988). A theory of rational addiction. *The Journal of Political Economy*, 675–700.

Becker, S. W., Ronen, J., & Sorter, G. H. (1974). Opportunity costs-an experimental approach. *Journal of Accounting Research*, 317–329.

Bentham, J. (1789/1979). *An introduction to the principles of morals and legislation*. T. Payne, and Son.

Bernoulli, D. (1738/1954). Exposition of a new theory on the measurement of risk. *Econometrica: Journal of the Econometric Society*, 23–36.

Binmore, K. (1999). Why experiment in economics? *The Economic Journal*, *109*(453), 16–24.

Bolton, G. E., & Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American economic review*, 166–193.

Braga, J., & Starmer, C. (2005). Preference anomalies, preference elicitation and the discovered preference hypothesis. *Environmental and Resource Economics*, *32*(1), 55–89.

Braun, O. L., & Wicklund, R. A. (1989). Psychological antecedents of conspicuous consumption. *Journal of Economic psychology*, *10*(2), 161–187.

Bruner, J. S. (1957). On perceptual readiness. *Psychological review*, *64*(2), 123.

Bruni, L., & Sugden, R. (2007). The road not taken: how psychology was removed from economics, and how it might be brought back*. *The Economic Journal*, *117*(516), 146–173.

Brunswig, A. (1910). *Das vergleichen und die relationserkenntnis*. Leibzig / Berlin: B. G. Teubner.

Buchanan, J. M. (2008). Opportunity cost. In S. N. Durlauf & L. E. Blume (Eds.), *The new palgrave dictionary of economics*. Basingstoke: Palgrave Macmillan.

Burnham, T. C. (2007). High-testosterone men reject low ultimatum game offers. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1623), 2327–2330.

Camerer, C. (2003a). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, C. (2003b). Strategizing in the brain. *Science*, *300*(5626), 1673–1675.

Camerer, C., & Thaler, R. (1995). Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives*, 209–219.

Chen, M. K., Lakshminarayanan, V., & Santos, L. R. (2006). How basic are behavioral biases? evidence from capuchin monkey trading behavior. *Journal of Political Economy*, *114*(3), 517–537.

Cox, J. C., & Grether, D. M. (1996). The preference reversal phenomenon: Response mode, markets and incentives. *Economic Theory*, *7*(3), 381–405.

Creedy, J. (2008). Edgeworth, francis ysidro (1845-1926). In S. N. Durlauf & L. E. Blume (Eds.), *The new palgrave dictionary of economics*. Basingstoke: Palgrave Macmillan.

Crockett, M. J., Clark, L., Tabibnia, G., Lieberman, M. D., & Robbins, T. W. (2008). Serotonin modulates behavioral reactions to unfairness. *Science*, *320*(5884), 1739–1739.

Dodds, W. B., Monroe, K. B., & Grewal, D. (1991). Effects of price, brand, and store information on buyers' product evaluations. *Journal of marketing research*, 307–319.

Easterlin, R. A. (1974). Does economic growth improve the human lot? some empirical evidence. *Nations and households in economic growth*, *89*, 89–125.

Edgeworth, F. Y. (1877). *New and old methods of ethics: or 'physical ethics' and 'methods of ethics'*. Parker, Oxford.

Edgeworth, F. Y. (1881). *Mathematical psychics: An essay on the application of mathematics to the moral sciences*. C. Keagann Paul, London.

Edwards, W. (1954). The theory of decision making. *Psychological bulletin*, *51*(4), 380–417.

Elder, R. S., & Krishna, A. (2012). The "visual depiction effect" in advertising: Facilitating embodied mental simulation through product orientation. *Journal of Consumer Research*, *38*(6), 988–1003.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, *54*(2), 293–315.

Fazio, R. H. (1986). How do attitudes guide behavior. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 204–243). New York: Guilford.

Fechner, G. T. (1860/1964). *Elemente der psychophysik*. Bonset, Amsterdam.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137–140.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*(3), 817–868.

Fisher, I. (1892/2007). *Mathematical investigations in the theory of value and prices*. Cosimo, New York.

Fiske, A. P. (1992). The four elementary forms of sociality: framework for a unified theory of social relations. *Psychological review*, *99*(4), 689.

Fiske, S. T. (1993). Social cognition and social perception. *Annual review of psychology*, *44*(1), 155–194.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic behavior*, *6*(3), 347–369.

Frederick, S., Novemsky, N., Wang, J., Dhar, R., & Nowlis, S. (2009). Opportunity cost neglect. *Journal of Consumer Research*, *36*(4), 553–561.

Friedman, L. A., & Neumann, B. R. (1980). The effects of opportunity costs on project investment decisions: A replication and extension. *Journal of Accounting Research*, 407–419.

Friedman, M. (1953). The methodology of positive economics. In *Essays in positive economics.* University of Chicago Press, Chicago.

Friedman, M., & Savage, L. J. (1948). The utility analysis of choices involving risk. *The Journal of Political Economy*, 279–304.

Friedman, M., & Savage, L. J. (1952). The expected-utility hypothesis and the measurability of utility. *The Journal of Political Economy*, 463–474.

Fudenberg, D., & Levine, D. K. (2006). A dual-self model of impulse control. *The American Economic Review*, 1449–1476.

Gilbert, D. T., Gill, M. J., & Wilson, T. D. (2002). The future is now: Temporal correction in affective forecasting. *Organizational Behavior and Human Decision Processes*, *88*(1), 430–444.

Glimcher, P. W., & Fehr, E. (2013). *Neuroeconomics: Decision making and the brain*. Academic Press.

Gossen, H. H. (1854). *Entwickelung der gesetze des menschlichen verkehrs, und der daraus fliessenden regeln für menschliche handeln*. Braunschweig: F. Vieweg & Sohn.

Green, D. I. (1894). Pain-cost and opportunity-cost. *The Quarterly Journal of Economics*, *8*(2), 218–229.

Grether, D. M., & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *The American Economic Review*, 623–638.

Gu, J., Bohns, V. K., & Leonardelli, G. J. (2013). Regulatory focus and interdependent economic decision-making. *Journal of Experimental Social Psychology*, *49*(4), 692–698.

Güth, W. (1995). On ultimatum bargaining experiments - a personal review. *Journal of Economic Behavior & Organization*, *27*(3), 329–344.

Güth, W., Huck, S., & Müller, W. (2001). The relevance of equal splits in ultimatum games. *Games and Economic Behavior*, *37*(1), 161–169.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization*, *3*(4), 367–388.

Handgraaf, M. J., Dijk, E. v., Wilke, H. A., & Vermunt, R. C. (2004). Evaluability of outcomes in ultimatum bargaining. *Organizational Behavior and Human Decision Processes*, *95*(1), 97–106.

Helson, H. (1964). *Adaptation-level theory*. Harper & Row.

Hewig, J., Kretschmer, N., Trippe, R. H., Hecht, H., Coles, M. G., Holroyd, C. B., & Miltner, W. H. (2011). Why humans deviate from rational choice. *Psychophysiology*, *48*(4), 507–514.

Heyman, J., & Ariely, D. (2004). Effort for payment a tale of two markets. *Psychological Science*, *15*(11), 787–793.

Hicks, J. R., & Allen, R. G. (1934a). A reconsideration of the theory of value. part i. *Economica*, *1*(1), 52–76.

Hicks, J. R., & Allen, R. G. (1934b). A reconsideration of the theory of value. part ii. a mathematical theory of individual demand functions. *Economica*, *1*(2), 196–219.

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and self-control from a dual-systems perspective. *Perspectives on Psychological Science*, *4*(2), 162–176.

Hofmann, W., Rauch, W., & Gawronski, B. (2007). And deplete us not into temptation: Automatic attitudes, dietary restraint, and self-regulatory resources as determinants of eating behavior. *Journal of Experimental Social Psychology*, *43*(3), 497–504.

Hoskin, R. E. (1983). Opportunity cost and behavior. *Journal of Accounting Research*, 78–95.

Houthakker, H. S. (1950). Revealed preference and the utility function. *Economica*, 159–174.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, *67*(3), 247–257.

Hsee, C. K. (1998). Less is better: When low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making*, *11*(2), 107–121.

Hsee, C. K. (1999). Value seeking and prediction-decision inconsistency: Why don't people take what they predict they'll like the most? *Psychonomic Bulletin and Review*, *6*(4), 555–561.

Hsee, C. K., & Hastie, R. (2006). Decision and experience: why don't we choose what makes us happy? *Trends in cognitive sciences*, *10*(1), 31–37.

Hsee, C. K., & Rottenstreich, Y. (2004). Music, pandas, and muggers: on the affective psychology of value. *Journal of Experimental Psychology: General*, *133*(1), 23.

Hsee, C. K., Yu, F., Zhang, J., & Zhang, Y. (2003). Medium maximization. *Journal of Consumer Research*, *30*, 1–14.

Hsee, C. K., & Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science*, *5*(4), 343–355.

Hsee, C. K., Zhang, J., Yu, F., & Xi, Y. (2003). Lay rationalism and inconsistency between predicted experience and decision. *Journal of Behavioral Decision Making*, *16*(4), 257–272.

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.

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology*, *79*(6), 995.

Jensen, K., Call, J., & Tomasello, M. (2007). Chimpanzees are rational maximizers in an ultimatum game. *Science*, *318*(5847), 107–109.

Jevons, W. S. (1871). *The theory of political economy*. Baltimore: Macmillan and Co.

Jevons, W. S. (1875). *Money and the mechanism of exchange*. New York: D. Appleton and Co.

Kagel, J. H., Battalio, R. C., Rachlin, H., Green, L., Basmann, R. L., & Klemm, W. R. (1975). Experimental studies of consumer demand behavior using laboratory animals. *Economic Inquiry*, *13*(1), 22–38.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological science*, *4*(6), 401–405.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review*, *76*(4), 728–741.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1990). Experimental tests of the endowment effect and the coase theorem. *Journal of political Economy*, 1325–1348.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological review*, *93*(2), 136.

Kahneman, D., Ritov, I., & Schkade, D. (1999). Economic preferences or attitude expressions?: An analysis of dollar responses to public issues. *Journal of Risk and Uncertainty*, *19*(13), 203–235.

Kahneman, D., & Snell, J. (1992). Predicting a changing taste: Do people know what they will like? *Journal of Behavioral Decision Making*, *5*(3), 187–200.

Kahneman, D., & Thaler, R. (2006). Anomalies: Utility maximization and experienced utility. *The Journal of Economic Perspectives*, *20*(1), 221–234.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292.

Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to bentham? explorations of experienced utility. *The Quarterly Journal of Economics*, 375–405.

Kalyanaram, G., & Winer, R. S. (1995). Empirical generalizations from reference price research. *Marketing Science*, *14*(3, Part 2 of 2), G161–G169.

Katzner, D. W. (2008). integrability of demand. In S. N. Durlauf & L. E. Blume (Eds.), *The new palgrave dictionary of economics.* Basingstoke: Palgrave Macmillan.

Keynes, J. M. (1939/2006). *General theory of employment, interest and money*. Atlantic Publishers & Dist.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*(5800), 829–832.

Knutson, B., Rick, S., Wimmer, G. E., Prelec, D., & Loewenstein, G. (2007). Neural predictors of purchases. *Neuron*, *53*(1), 147–156.

Koenigs, M., & Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: evidence from the ultimatum game. *The Journal of neuroscience*, *27*(4), 951–956.

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

Larrick, R. P., Morgan, J. N., & Nisbett, R. E. (1990). Teaching the use of cost-benefit reasoning in everyday life. *Psychological Science*, *1*(6), 362–370.

Larrick, R. P., Nisbett, R. E., & Morgan, J. N. (1993). Who uses the cost-benefit rules of choice? implications for the normative status of microeconomic theory. *Organizational Behavior and Human Decision Processes*, *56*(3), 331–347.

Levine, D. K. (2012). *Is behavioral economics doomed?: The ordinary versus the extraordinary*. Cambridge: Open Book Publishers.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of experimental psychology*, *89*(1), 46.

Lichtenstein, S., & Slovic, P. (2006). *The construction of preference* (S. Lichtenstein & P. Slovic, Eds.). Cambridge University Press.

Lo, A. W. (2008). efficient markets hypothesis. In S. N. Durlauf & L. E. Blume (Eds.), *The new palgrave dictionary of economics*. Basingstoke: Palgrave Macmillan.

Lockwood, B. (2008). Pareto efficiency. In S. N. Durlauf & L. E. Blume (Eds.), *The new pal-*

grave dictionary of economics. Basingstoke: Palgrave Macmillan.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational behavior and human decision processes*, *65*(3), 272–292.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American economic review*, 426–432.

Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision making. In *Handbook of affective sciences* (pp. 619–642). Oxford University Press.

Loewenstein, G., Nagin, D., & Paternoster, R. (1997). The effect of sexual arousal on expectations of sexual forcefulness. *Journal of Research in Crime and Delinquency*, *34*(4), 443–473.

Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Projection bias in predicting future utility. *The Quarterly Journal of Economics*, *118*(4), 1209–1248.

Loewenstein, G., & Schkade, D. (1999). Wouldn't it be nice? predicting future feelings. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 85–105). New York: Russell Sage Foundation Press.

Loomes, G., Starmer, C., & Sugden, R. (2010). Preference reversals and disparities between willingness to pay and willingness to accept in repeated markets. *Journal of Economic Psychology*, *31*(3), 374–387.

Machina, M. J. (2008). expected utility hypothesis. In S. N. Durlauf & L. E. Blume (Eds.), *The new palgrave dictionary of economics.* Basingstoke: Palgrave Macmillan.

Malinvaud, E. (1952). Note on von neumann-morgenstern's strong independence axiom. *Econometrica: Journal of the Econometric Society*, 679–679.

Markus, H. R., & Schwartz, B. (2010). Does choice mean freedom and well-being? *Journal of Consumer Research*, *37*(2), 344–355.

Marschak, J. (1950). Rational behavior, uncertain prospects, and measurable utility. *Econometrica: Journal of the Econometric Society*, 111–141.

Mayhew, G. E., & Winer, R. S. (1992). An empirical analysis of internal and external reference prices using scanner data. *Journal of consumer Research*, 62–70.

Menger, C. (1871). *Grundsätze der volkswirtschaftslehre*. Wilhelm Braumüller.

Menger, C. (1892). On the origin of money. *The Economic Journal*, *2*(6), 239–255.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual review of psychology*, *32*(1), 89–115.

Mill, J. S. (1843). *A system of logic, ratiocinative and inductive (vol. i & ii)*. John W. Parker, London.

Mill, J. S. (1848). *Principles of political economy (vol. i & ii)*. John W. Parker, London.

Monroe, K. B. (1973). Buyers' subjective perceptions of price. *Journal of marketing research*, 70–80.

Morewedge, C. K., Krishnamurti, T., & Ariely, D. (2014). Focused on fairness: Alcohol intoxication increases the costly rejection of inequitable rewards. *Journal of Experimental Social Psychology*, *50*, 15–20.

Morgenstern, O. (1979). Some reflections on utility. In M. Allais & O. Hagen (Eds.), *Expected utility hypothesis and the allais paradox.* Reidel, Dordrecht.

Mussel, P., Göritz, A. S., & Hewig, J. (2013). Which choice is the rational one? an investigation of need for cognition in the ultimatum game. *Journal of Research in Personality*, *47*(5), 588–591.

Mussweiler, T., & Epstude, K. (2009). Relatively fast! efficiency advantages of comparative thinking. *Journal of Experimental Psychology: General*, *138*(1), 1.

Mussweiler, T., & Strack, F. (1999). Comparing is believing: A selective accessibility model of judgmental anchoring. *European review of social psychology*, *10*(1), 135–167.

Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin*, *26*(9), 1142–1150.

Neumann, B. R., & Friedman, L. A. (1978). Opportunity costs: Further evidence through an experimental replication. *Journal of Accounting Research*, 400–410.

Niehans, J. (1990). *A history of economic theory: Classic contributions 1720-1980*. Baltimore: John Hopkins University Press.

Nowak, M. A., Page, K. M., & Sigmund, K. (2000). Fairness versus reason in the ultimatum game. *Science*, *289*(5485), 1773–1775.

Pareto, V. (1909/1971). *Manual of political economy*. Kelley, New York. (English translation of the

1909 French edition of the 1906 Italian Manuale d'economia politica con una introduzione alla scienza sociale, Milan: Societa Editrice Libraria.)

Pham, M. T. (2007). Emotion and rationality: A critical review and interpretation of empirical evidence. *Review of General Psychology*, *11*(2), 155.

Plott, C. R. (1996). Rational individual behavior in markets and social choice processes: the discovered preference hypothesis. In K. Arrow, E. Colombatto, M. Perleman, & C. Schmidt (Eds.), *Rational foundations of economic behavior* (pp. 225–250). Macmillan and St. Martin's, London.

Prelec, D., & Loewenstein, G. (1998). The red and the black: Mental accounting of savings and debt. *Marketing Science*, *17*(1), 4–28.

Raghubir, P., & Srivastava, J. (2008). Monopoly money: the effect of payment coupling and form on spending behavior. *Journal of Experimental Psychology: Applied*, *14*(3), 213.

Raghubir, P., & Srivastava, J. (2009). The denomination effect. *Journal of Consumer Research*, *36*(4), 701–713.

Rao, A. R., & Monroe, K. B. (1989). The effect of price, brand name, and store name on buyers' perceptions of product quality: an integrative review. *Journal of marketing Research*, 351–357.

Ratneshwar, S., Shocker, A. D., & Stewart, D. W. (1987). Toward understanding the attraction effect: The implications of product stimulus meaningfulness and familiarity. *Journal of Consumer Research*, *13*(4), 520–33.

Rawls, J., Freeman, S. R., & Schulte, J. (2008). *Geschichte der politischen philosophie*. Frankfurt: Suhrkamp.

Read, D. (2007). Experienced utility: utility theory from jeremy bentham to daniel kahneman. *Thinking & Reasoning*, *13*(1), 45–61.

Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: a randomized trial. *Pain*, *104*(1), 187–194.

Rosch, E. (1973). Natural categories. *Cognitive psychology*, *4*(3), 328–350.

Rosch, E., & Lloyd, B. B. (1978). *Cognition and categorization*. Hillsdale, New Jersey.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, 61–71.

Samuelson, P. A. (1950). The problem of integrability in utility theory. *Economica*, 355–385.

Samuelson, P. A. (1952). Probability, utility, and the independence axiom. *Econometrica: Journal of the Econometric Society*, 670–678.

Samuelson, P. A. (1953). Consumption theorems in terms of overcompensation rather than indifference comparisons. *Economica*, 1–9.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*(5626), 1755–1758.

Savage, L. J. (1954). *The foundations of statistics*. Courier Dover Publications.

Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? a meta-analytic review of choice overload. *Journal of Consumer Research*, *37*(3), 409–425.

Schindler, R. M. (1989). The excitement of getting a bargain: some hypotheses concerning the origins and effects of smart-shopper feelings. *Advances in consumer research*, *16*(1), 447–453.

Schindler, R. M. (1998). Consequences of perceiving oneself as responsible for obtaining a discount: evidence for smart-shopper feelings. *Journal of Consumer Psychology*, *7*(4), 371–392.

Schumpeter, J. A. (1954/2009). *Geschichte der ökonomischen analyse ii* (E. B. Schumpeter, Ed.). Vandenhoeck & Ruprecht, Göttingen.

Schwarz, N. (2011). Feelings-as-information theory. *Handbook of theories of social psychology*, *1*, 289–308.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social psychology*, *61*(2), 195.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of personality and social psychology*, *45*(3), 513.

Schwarz, N., Strack, F., & Mai, H.-P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public opinion quarterly*, *55*(1), 3–23.

Sedikides, C., Ariely, D., & Olsen, N. (1999). Contextual and procedural determinants of partner selection: Of asymmetric dominance and prominence. *Social Cognition*, *17*(2), 118–139.

Sen, A. (1973). Behaviour and the concept of preference. *Economica*, 241–259.

Sen, A. (2008). rational behaviour. In S. N. Durlauf & L. E. Blume (Eds.), *The new palgrave dictionary of economics*. Basingstoke: Palgrave Macmillan.

Shafir, E., & Thaler, R. (2006). Invest now, drink later, spend never: On the mental accounting of delayed consumption. *Journal of economic psychology*, *27*(5), 694–712.

Shiller, R. J. (2015). *Irrational exuberance*. Princeton University Press.

Shiv, B., Carmon, Z., & Ariely, D. (2005). Placebo effects of marketing actions: Consumers may get what they pay for. *Journal of marketing Research*, *42*(4), 383–393.

Shiv, B., & Huber, J. (2000). The impact of anticipating satisfaction on consumer choice. *Journal of Consumer Research*, *27*(2), 202–216.

Shogren, J. F., Shin, S. Y., Hayes, D. J., & Kliebenstein, J. B. (1994). Resolving differences in willingness to pay and willingness to accept. *The American Economic Review*, 255–270.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118.

Simon, H. A. (1959). Theories of decision-making in economics and behavioral science. *The American economic review*, 253–283.

Simon, H. A. (1983). *Reason in human affairs*. Stanford University Press, Redwood City.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*(2), 158–74.

Simonson, I. (2008). Will i like a "medium" pillow? another look at constructed and inherent preferences. *Journal of Consumer Psychology*, *18*(3), 155–169.

Simonson, I., & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of marketing research*.

Slesnick, D. T. (2008). consumer surplus. In S. N. Durlauf & L. E. Blume (Eds.), *The new palgrave dictionary of economics.* Basingstoke: Palgrave Macmillan.

Slovic, P. (1995). The construction of preference. *American psychologist*, *50*(5), 364.

Slovic, P., Griffin, D., & Tversky, A. (1990). Compatibility effects in judgment and choice. *Insights in decision making: A tribute to Hillel J. Einhorn*, 5–27.

Slutsky, E. (1915/1953). On the theory of the budget of the consumer. In *Readings in price theory.* Allen & Unwin.

Smith, V. L. (1994). Economics in the laboratory. *The Journal of Economic Perspectives*, 113–131.

Spiller, S. A. (2011). Opportunity cost consideration. *Journal of Consumer Research*, *38*(4), 595–610.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of economic literature*, 332–382.

Stigler, G. J. (1950a). The development of utility theory. i. *The Journal of Political Economy*, 307–327.

Stigler, G. J. (1950b). The development of utility theory. ii. *The Journal of Political Economy*, 373–396.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and social psychology review*, *8*(3), 220–247.

Strack, F., & Martin, L. L. (1987). Thinking, judging, and communicating: A process account of context effects in attitude surveys. In *Social information processing and survey methodology* (pp. 123–148). Springer.

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, *73*(3), 437–446.

Strack, F., Schwarz, N., & Gschneidinger, E. (1985). Happiness and reminiscing: The role of time perspective, affect, and mode of thinking. *Journal of Personality and Social Psychology*, *49*(6), 1460.

Sugden, R. (1991). Rational choice: A survey of contributions from economics and philosophy. *Economic Journal*, *101*(407), 751–785.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, *1*(1), 39–60.

Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economic Letters*, *8*, 201–207.

Thaler, R. (1983). Transaction utility theory. *Advances in Consumer Research*, *10*.

Thaler, R. (1985/2008). Mental accounting and consumer choice. *Marketing Science*, *27*(1), 15–25.

Thaler, R. (1988). The ultimatum game. *The Journal of Economic Perspectives*, *2*(4), 195–206.

Thaler, R. (1994). *The winner's curse: Paradoxes and anomalies of economic life*. Princeton, New Jersey: Princeton University Press.

Thaler, R. (1999). Mental accounting matters. *Journal of Behavioral decision making*, *12*(3), 183–206.

Thaler, R., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Thomas, M., Desai, K. K., & Seenivasan, S. (2011). How credit card payments increase unhealthy food purchases: visceral regulation of vices. *Journal of consumer research*, *38*(1), 126–139.

Tversky, A., & Griffin, D. (1991). Endowment and contrast in judgments of well-being. In F. Strack, M. Argyle, & N. Schwarz (Eds.), *Subjective well-being: an interdisciplinary perspective* (Vol. 21, pp. 101–118). Pergamon Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124–1131.

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

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297–323.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological review*, *95*(3), 371.

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, *39*(10), 1179–1189.

Tversky, A., & Thaler, R. (1990). Anomalies: Preference reversals. *Journal of Economic Perspectives*, *4*(2), 201–211. (Spring)

Van Boven, L., Dunning, D., & Loewenstein, G. (2000). Egocentric empathy gaps between owners and buyers: misperceptions of the endowment effect. *Journal of personality and social psychology*, *79*(1), 66.

van't Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2006). Affective state and decision-making in the ultimatum game. *Experimental Brain Research*, *169*(4), 564–568.

Varian, H. R. (1982). The nonparametric approach to demand analysis. *Econometrica: Journal of the Econometric Society*, 945–973.

Varian, H. R. (2006). Revealed preference. In M. Szenberg, L. B. Ramrattan, & A. A. Gottesman (Eds.), *Samuelsonian economics and the 21st century.* Oxford University Press,.

von Neumann, J., & Morgenstern, O. (1944/1955). *Theory of games and economic behavior* (Third Edition ed.). Princeton University Press.

von Wieser, F. (1889). *Der natürliche werth*. Hölder, Wien.

von Wieser, F. (1914/1924). *Theorie der gesellschaftlichen wirtschaft*. Paul Siebeck, Tübingen.

Walras, L. (1874). *Éléments d'économie politique pure; ou, Théorie de la richesse sociale*. L. Corbaz & cie.

Weber, E. H. (1846/2006). *Tastsinn und gemeingefühl*. VDM Müller, Saarbrücken. ([Nachdr. aus:] Handwörterbuch der Physiologie, [Bd. 3, Braunschweig, Vieweg], 1846)

Weg, E., & Zwick, R. (1994). Toward the settlement of the fairness issues in ultimatum games: A bargaining approach. *Journal of Economic Behavior & Organization*, *24*(1), 19–34.

Winer, R. S. (1986). A reference price model of brand choice for frequently purchased products. *Journal of consumer research*, 250–256.

Wundt, W. (1874). *Grundzüge der physiologischen psychologie*. W. Engelmann, Leibzig.

Yang, X. (2001). *Economics: New classic versus neoclassical frameworks*. Blackwell, Massachusetts.

Zauberman, G., & Lynch Jr, J. G. (2005). Resource slack and propensity to discount delayed investments of time versus money. *Journal of Experimental Psychology: General*, *134*(1), 23.

Appendices

Study 1

After agreeing to participate in the study, subjects were redirected to an external site (*EFS survey*) were the following introduction was displayed.

> Dear Participant,
> in the following survey you will first of all face some daily life situations, in which someone needs a helping hand. You will be asked to estimate the average person's willingness to help in each of the situations. The rating scale will reach from 1, not at all likely to help, to 11, will help for sure, with proportionate categories in between. Please tick the category that most likely matches your assessment. There is no right or wrong in this task. Please answer the questions to the best of your judgment. After this task you will be asked some questions on your person, on situations and items. All the data will be collected anonymously and you will have the possibility to cancel the questionnaire at any time.

Scenarios

Then, the following five scenarios were presented in a randomized order.

Car

Imagine you are driving on a barely frequented road and encounter a broken-down car. You stop and the driver asks you to lend him your tool kit. How would you rate **an average person's** likelihood to help if the driver offered an invitation for...

Beach

Imagine you are at a quite crowded beach. The couple lying next to you asks you if you would mind watching their bags and towels while they go swimming in order to protect them from theft. How would you rate **an average person's** likelihood to help if being offered...

Sofa

Imagine you are walking through the town you live in without a specific purpose or appointment in mind. You happen to pass some people who are currently moving house. One of the men steps out and politely asks you to help them load a sofa into a van. How would you rate **an average person's** likelihood to help if being offered...

Accident

> Imagine you get a call from a woman that just moved to your town. You got to know her at a parents' evening. One of her children fell and she needs to take him/her to the hospital to get the cut stitched. Her other children will be home from school soon and she has no one to take care of them until she is back. She asks you if you could help her out. How would you rate **an average person's** likelihood to help if the neighbor offered...

Cemetery

> Imagine you are at a cemetery where an older lady is taking care of her husband's grave. You get into a conversation with her and she tells you that she wants to visit her daughter abroad for 2 weeks. She has no one to water the flowers and, as she has noticed that you come there regularly, she asks you to help her. How would you rate **an average person's** likelihood to help if the lady offered...

Stimuli

The first study included five scenarios for which participants indicated their anticipated effort levels depending on different forms and levels of payment. The pay-

ments for each scenario, payment form and payment level are depicted in Table A1.

In addition, the design in Study 1 varied the evaluation mode in which judgments about anticipated effort levels were made. Specifically, participants either evaluated both payment levels (joint evaluation) or only one (separate evaluation). Figure A1 illustrates the task in both conditions.

Plausibility Check

Several items were included to test for different evaluations of the payments and the tasks across conditions. Table A2 shows the price estimates for the non-monetary payments used in this study. Please note that the *p*-value in the last column refers to an ANOVA testing for differences across conditions.

Furthermore, participants were asked to rate the anticipated strain and duration of the tasks (see Table A3). Please note that the *p*-value in the last column refers to an ANOVA testing for differences across conditions.

Payment Form	Scenario	Payment Level	
		Low	High
	Car		
Monetary		$1.50	$15
Non-Monetary		Coffee	Lunch
	Beach		
Monetary		$1	$8
Non-Monetary		Beer	Cocktail
	Sofa		
Monetary		$0.50	$5
Non-Monetary		Candy Bar	Chocolate Box
	Accident		
Monetary		$2	$20
Non-Monetary		Cookie	Cake
	Cemetery		
Monetary		$5	$50
Non-Monetary		Mini Bottle of Wine	Wine Selection

Table A1: Overview of payments in all five scenarios.

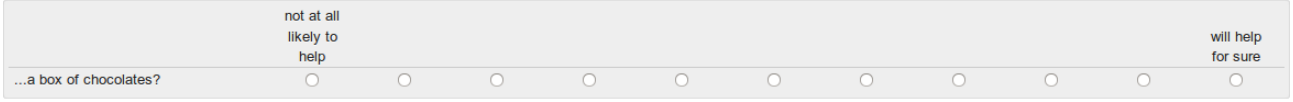
Product	Non-Monetary			Monetary	
	Low	High	Joint	Joint	p-value
Cup of coffee	2.21	2.38	4.16	2.16	0.49
Lunch	8.67	8.90	14.23	9.16	0.18
Beer	4.21	4.38	5.29	3.71	0.75
Cocktail	6.14	5.90	7.27	6.38	0.84
Candybar	1.24	1.27	2.53	1.45	0.42
Box of Chocolates	8.26	8.47	11.61	8.72	0.63
Cookie	1.32	1.55	1.97	1.56	0.38
Cake	8.20	9.85	17.79	10.69	0.37
Miniwine	4.50	5.42	7.49	5.57	0.25
Selection of wine	14.00	17.56	33.98	17.50	0.13

Table A2: Price estimates (in $) for the ten non-monetary payments.

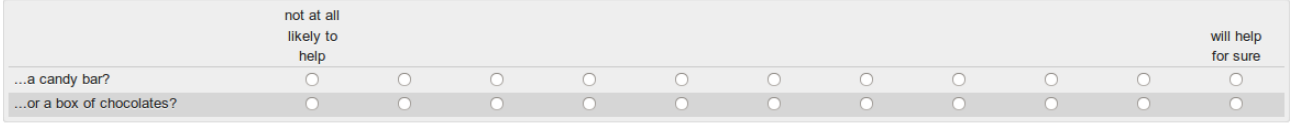
Product		Non-Monetary Low	Non-Monetary High	Non-Monetary Joint	Monetary Joint	p-value
Car						
	Strain	1.32	1.46	1.33	1.37	0.72
	Duration	37	34	37	46	0.16
Beach						
	Strain	1.32	1.36	1.29	1.27	0.91
	Duration	31	29	37	31	0.52
Sofa						
	Strain	3.30	3.46	3.39	3.31	0.88
	Duration	12	12	14	15	0.54
Accident						
	Strain	3.08	2.92	3.18	3.18	0.67
	Duration	166	158	183	188	0.52
Cementary						
	Strain	2.04	1.72	1.73	1.86	0.33
	Duration	218	80	89	88	0.24

Table A3: Ratings about expected strain (1 = low; 5 = high) and duration (in minutes) of the tasks described in each scenario.

A



B



C

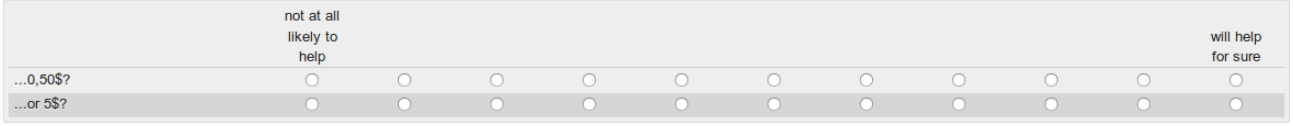


Figure A1: Part A of the figure illustrates the task in single evaluation mode for non-monetary payments. Part B shows the joint evaluation mode for non-monetary payments. Part C depicts the joint evaluation mode for monetary payments.

Study 2

Materials

On the external site, participants first read the following introduction:

> Welcome to our study.
> This study concerns hypothetical situations and will only take a few minutes. Please read the scenarios carefully and try to anticipate how you would act if you actually were in this situation.
> Of course there are no right and no wrong answers.
> Click Continue to begin with the study.

Afterwards, the following scenario was presented:

> Please try to imagine the following situation:
> One day you wake up and decide that you want to have an aquarium. You go out and get an appropriate tank, plants and some other supplies.
> The only task left now is to decide which fish to put into the aquarium.
> In the shop they have two kinds of fish that would fit your tank and you have **$20** left. For the aquarium to look nice you should spend the whole $20.
> Click on Continue to find out which fish are available.

Participants then indicated an allocation of the $20 (see Figure A2).

On the next screen, they were presented the second part of the scenario.

> The aquarium looks very nice and you like to watch the fish once in a while.
> Some time later you feel that an aquarium would also look nice in the other room so you quickly decide to get another one. Again, after getting all the other things, you have to decide which fish to buy for the other aquarium.
> You want to spend another **$20** for the new fish but in the shop you notice that **the prices of the fish have changed.**
> Click on Continue to find out which fish are available.

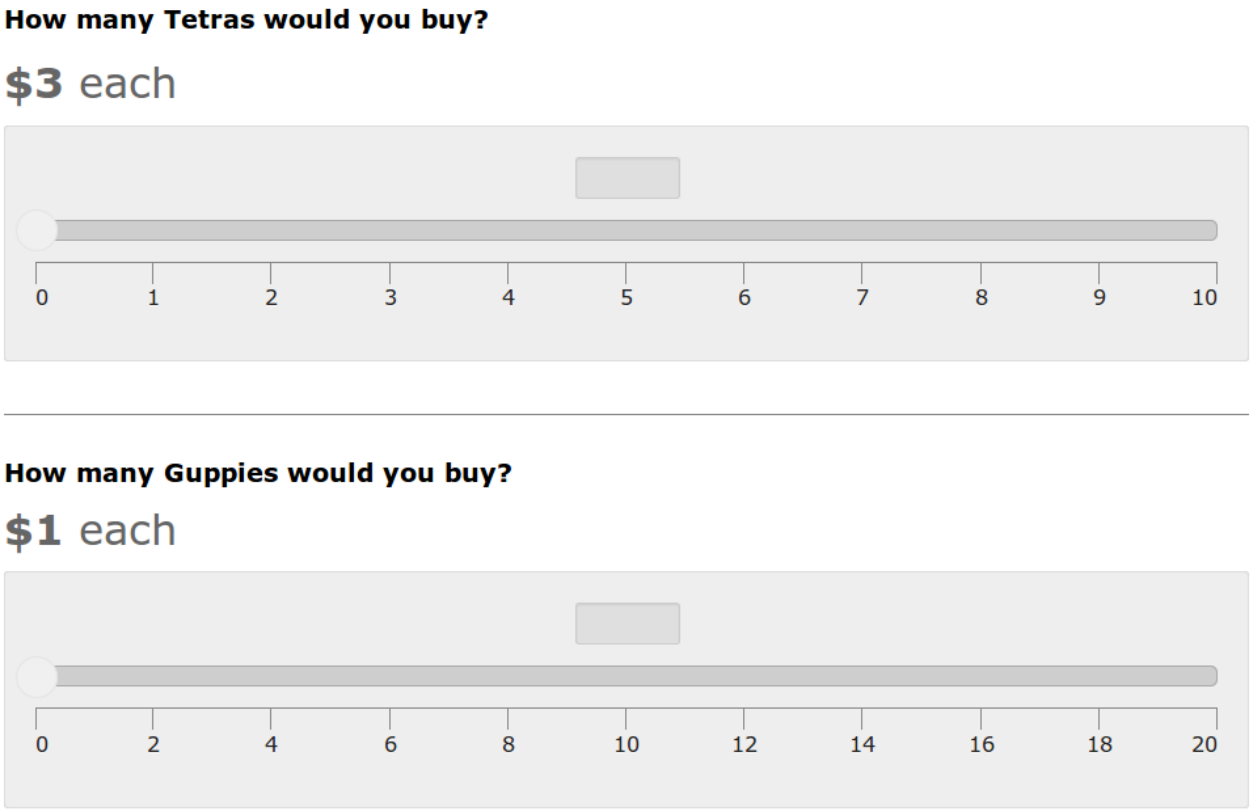


Figure A2: The figure illustrates how participants indicated their allocation in Study 2.

Study 3

Instructions

On the external site, participants first read the following introduction:

> Thank you for your willingness to take part in this questionnaire. The questionnaire is for scientific purposes only and has no commercial goals. Your data will be treated anonymously.
> The questionnaire will take about 15 minutes to complete.
> After you have finished the questionnaire, you will receive a code to enter at the mTurk HIT.
> Please click on Continue to start the questionnaire.

The next screen included the coverstory:

> This study is about your behavior in hypothetical situations. Specifically, please imagine you are shopping in the mall nearby.
> You are not after anything specific, but you browse through some stores and some of the products catch your attention. The products will be presented to you in a moment and you will have to answer some questions about them.
> Please try to imagine that you are in fact faced with the decisions

```
described and to imagine the situation
vividly.
Please click on Continue to see the
first product.
```

Materials

The third study investigated purchase decisions concerning twelve consumer products. Importantly, for each product, either a low or a high anchor was randomly selected before price estimates were assessed. The actual sale prices presented afterwards were the same for all participants (see Table A4).

Product	*Low Anchor*	*High Anchor*	Sale Price
armchair	100	500	300
ball	5	30	15
coin	20	160	90
DVD-player	15	125	70
flatscreen TV	200	1300	750
fridge	100	500	300
laptop	200	1100	650
laptop bag	10	70	40
suitcase	30	210	120
toothbrush	20	100	60
wall clock	10	60	35
washingmachine	200	1200	700

Table A4: Overview of anchors and sale prices for each product in Study 3.

Anchoring Effect

The anchor significantly affected Δp for all products except for the refrigerator (see Table A5). Nonetheless, the

refrigerator was not excluded for the following mediation analysis.

Product	*Low Anchor*	*High Anchor*	p-value
armchair	356	534	< 0.001
ball	7	16	< 0.001
coin	31	104	< 0.001
DVD-player	59	99	< 0.001
flatscreen TV	436	791	< 0.001
fridge	582	652	0.067
laptop	487	748	< 0.001
laptop bag	40	60	< 0.001
suitcase	74	150	< 0.001
toothbrush	37	63	< 0.001
wall clock	26	39	< 0.001
washingmachine	504	829	< 0.001

Table A5: Price estimates in the low and high anchor conditions in Study 3.

The relative price difference Δp was identified as a mediator for the evaluation of transactions. Figure A3 shows Δp for each product and anchor. The p-values refer to t-tests.

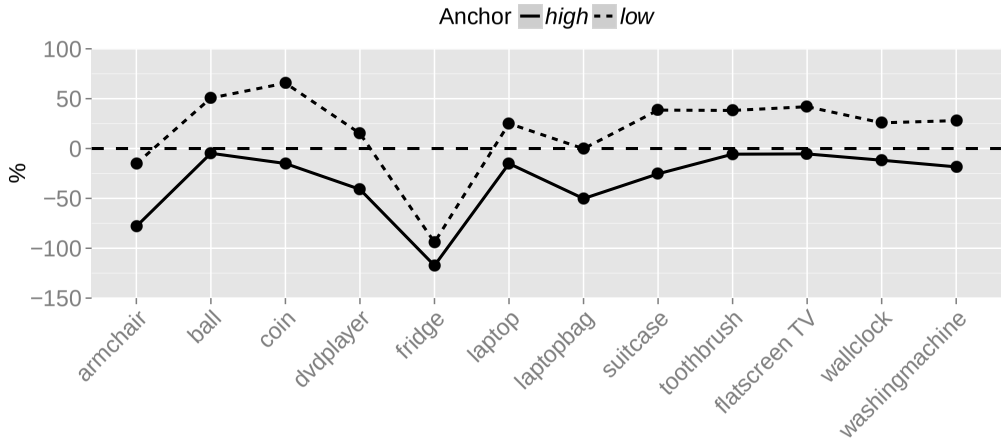


Figure A3: The figure illustrates Δp for all products depending on the preceding anchor. Again, the smooth line (high anchors) is below the dotted line (low anchors) for all products indicating a successful manipulation of Δp.

Study 4

Materials

Similar to Study 3, participants evaluated twelve transactions with varying anchors for reference price estimates. Table A6 shows both anchors for each product as well as the respective sale prices.

Product	*Low Anchor*	*High Anchor*	Sale Price
armchair	300	600	450
ball	5	45	25
coin	20	140	70
DVD-player	30	120	80
flatscreen TV	200	1100	650
fridge	500	800	650
laptop	200	1100	600
laptop bag	20	80	50
suitcase	30	210	120
toothbrush	20	80	50
wall clock	10	60	35
washingmachine	200	1200	650

Table A6: Overview of anchors and sale prices for each product in Study 4.

In addition, the payment mode varied between participants. In the credit card condition, participants selected which card to use (see Figure A4A). In contrast, participants in the cash condition selected the denomination of their payment (see Figure A4B).

Anchoring Effect

The results show strong anchoring effects for all products ($ps > 0.05$) except for the refrigerator ($t(111.83) = 1.26$, $p = 0.21$). However, the refrigerator was not excluded from the further analysis because the direct

Figure A4: Part A shows the options in the credit card condition. Part B shows the options from the cash condition where participants had to select which bills they wanted to use for payment. Necessarily, the available options differed across products in the cash condition.

Product	*Low Anchor*	*High Anchor*	*p*-values
armchair	442	546	0.011
ball	7	17	< 0.001
coin	46	104	< 0.001
DVD-player	52	94	< 0.001
flatscreen TV	497	740	< 0.001
fridge	758	821	0.212
laptop	464	722	< 0.001
laptop bag	38	48	0.001
suitcase	67	143	< 0.001
toothbrush	34	57	< 0.001
wall clock	21	36	< 0.001
washingmachine	461	763	< 0.001

Table A7: Price estimates in the low and high anchor conditions in Study 4.

mechanism underlying comparative utility was tested which is assumed to hinge on Δp rather than on anchor itself. Figure A5 shows Δp for each product and anchor.

The relative price difference Δp was again identified as a mediator for the evaluation of transactions. Figure A5 shows Δp for each product and anchor. The p-values refer to t-tests.

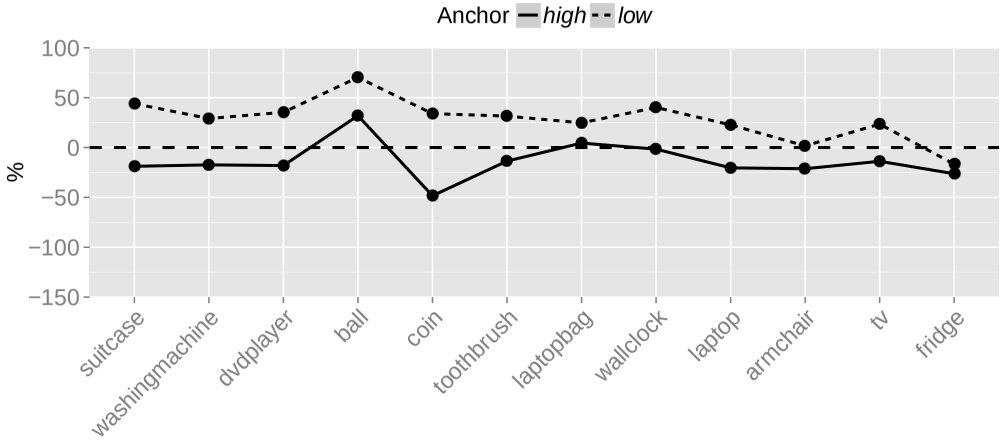


Figure A5: The figure shows Δp across products depending on the preceding anchor. Even though Δp varies across products, the smooth line (high anchors) is below the dotted line (low anchors) for all products indicating a successful manipulation of Δp through the anchors.

Study 5

Instructions

The following instructions explained the Ultimatum Game and also include the cover story featuring different currencies.

Welcome!
In international commerce, cooperation between companies from different parts of the world is growing ever more common. The division of joint profits is often the result of protracted negotiations.
In this task, strategies of international negotiation are examined using a simple game. The main question is what influence the type of currency has on the division of monetary sums.
The negotiations take place in a greatly simplified manner, such that one party (the proposer) makes a suggestion as to how a certain monetary sum should be divided and the other party (the responder) can then either accept or decline this suggestion. If the offer is declined, nobody gets anything. If on the other hand it is accepted, each party receives their share.
You will negotiate with a randomly chosen participant in each round of this game. In total, you will perform 150 of these simplified

negotiations. Your negotiation partners will change constantly, your role will however remain the same. You will be assigned a role (proposer or responder) at the beginning and will retain this role throughout the entire experiment.
Please make your decisions according to what you think best, as your final payout is directly dependent on your decisions!
The currency in which the monetary sums that are to be divided is displayed will be different for each party. However, all sums will automatically be converted to your currency.

After reading the instructions and the coverstory participants were informed that they would actually receive 1% of their total earnings after completing the task. Subsequently, each participant was ostensibly assigned to a role and a currency at random, but in fact all participants received the same information that they would play in the role of a responder and would receive offers in Euro Cent:

You have the role of the RESPONDER.
All sums offered to you are presented in EURO-CENTS.

After summarizing the rules of the task once more, participants learned about response keys (counterbalanced) and played one training trial. Additionally, participants in the difficult-to-compare condition were reminded that odd amounts of money were the result of a currency conversion.

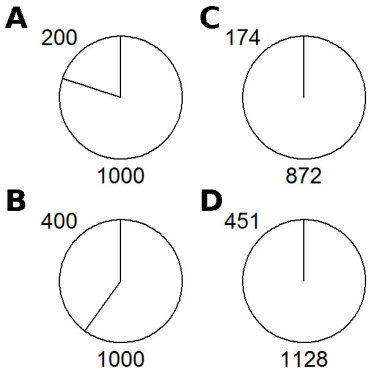


Figure A6: Ultimatum Game stimuli from Study 5. Part A shows a 20% offer from the easy-to-compare condition, Part B shows a 40% offer in the easy-to-compare condition. Part C shows a 20% offer from the difficult-to-compare condition, Part D shows a 40% offer in the difficult-to-compare condition.

Figure A6 shows a sample of stimuli from the ultimatum game across experimental conditions. To further decrease the typical ease of comparing the proposal to the cake in the first study offers in the difficult to compare condition were not illustrated by pie diagrams. After playing 150 trials, the total earnings were displayed and paid out.

Study 6

Instructions

Before playing the Ultimatum Game in Study 2, participants completed a short task to systematically highlight the consumptive consequences of money. Approximately two thirds of the participants were given the following instructions:

> Welcome to this experiment!
> Before we begin with the actual experiment, we would like to ask you to make a decision concerning different company brands.
> In this experiment, your reward will be paid out as a coupon. You now have exactly 2 minutes to think about what you could buy from these four shops. After 2 minutes, you will make your decision.

Participants were then shown the four store logos depicted in Part A of Figure A7. After two minutes of contemplation about possible purchases in these stores, participants indicated from which store they would like to receive a gift voucher:

> Please select the shop that you would most like to use your coupon in now. To do so, use the mouse and confirm your choice with the ENTER key or by clicking on the flashing button.
> Which shop would you most like to use your coupon in?

Figure A7: Stimuli from the task designed to highlight the consumptive consequences of money in Study 2. Part A shows the logos of stores displayed in both experimental conditions. Here, participants thought about possible purchases for 2 minutes. Part B shows city signs from the task in the control condition. Here, participants were asked to think about the sizes of the four cities for two minutes.

In the control condition, stores were replaced by cities (see Part B of Figure A7) and participants were asked to think about the city sizes for two minutes and then select the city with the most inhabitants. Specifically, the instructions were:

> Welcome to this experiment!
> Before we begin with the actual experiment, we would like to ask you to make a decision concerning different cities.
> You now have exactly 2 minutes to think about which of these cities has the most inhabitants. After 2 minutes, you will make your decision.
> Please select the city with the most inhabitants now. To do so, use the mouse and confirm your choice with the ENTER key or by clicking on the flashing button.
> Which city has the most inhabitants?

After making their choices (either about stores or cities) the instructions for the Ultimatum Game were

presented. The instructions were identical to those from Study 5 except that participants were informed that they would play 240 rounds and that participants in both experimental conditions were reminded that they would receive their 1% share of the earnings in the form of a gift voucher. The offer in the training trial was again identical in all conditions, however its presentation varied across experimental conditions. Figure A8 shows the different offer presentations across conditions from the actual game.

At the very end of the experiment, all participants received an envelope containing their earnings from the ultimatum game plus the show-up fee. Participants in the experimental conditions who were expecting a gift voucher received an envelope with the logo of the selected store and the words "Gutschein" (gift voucher) printed on.

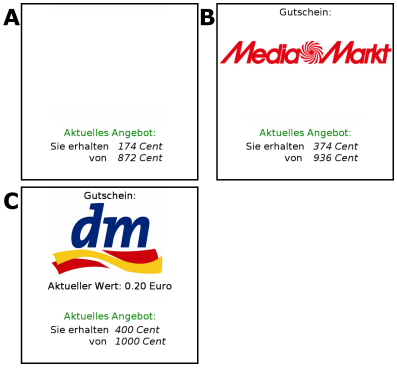


Figure A8: Ultimatum Game stimuli from Study 6. Part A shows a difficult-to-compare offer (20%) from the control condition where consumptive consequences were not highlighted. Part B shows a difficult-to-compare offer (40%) in the coupon condition. Part C shows an easy to compare offer (20%) in the coupon & sum condition. In both experimental conditions, the logo of the store from which participants ostensibly received a gift voucher was displayed with each offer.

Additional Results

Figure A9 shows the parabolas for both easy- and difficult-to-compare offers from Study 6.

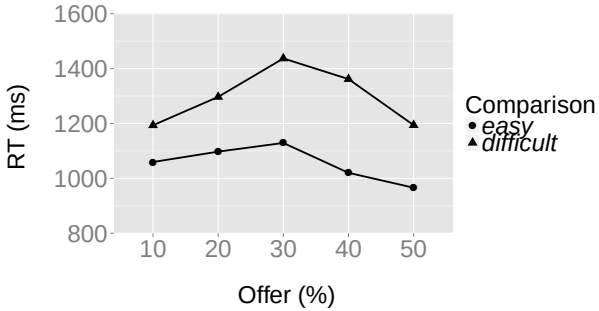


Figure A9: The figure shows the inverted U-shaped relationship between response latencies and relative offers in the easy and difficult to compare conditions. Similar to Study 5, latencies for difficult offers are generally longer and the parabola shape is more pronounced.

Additionally, Figure A10 illustrates the decreased acceptance gradient for difficult offers.

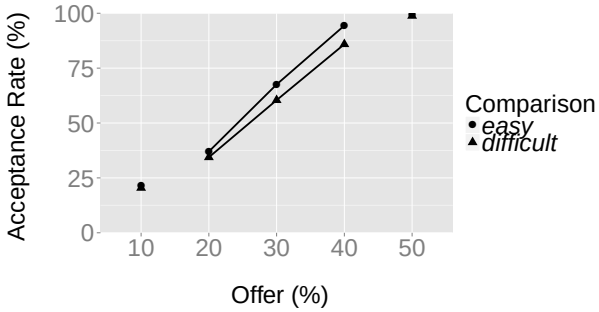


Figure A10: This figure shows the sigmoid relationship between acceptance rates and relative offers and the linear approximation for intermediate offers. Parallel to Study 5, the slope decreases if comparisons are difficult.