Stockholm School of Economics

Price and Concept as Complements or Substitutes: Signaling and Perception in Wine Tasting

Thesis within International Economics
(Course 5210, Curriculum 2000)

Original Version: 15 May 2013
Thesis Seminar: 30 May 2013
Final Version: 4 June 2013

Author:
Erik Andermo, 20946

Supervisor:
Assoc. Prof. Jesper Roine

Opponent(s):
Max Cederborg, 21791
Georg von Zweigbergk, 21605

Examinator:
Assoc. Prof. David Domeij

Keywords: Wine tasting, perception, expectations, price-quality heuristic, choice theory, neuroscience, incentives.
JEL Classification: C91, C93, D03, D83, D84, D87, L66, M31.

Abstract

Previous research has established a correlation between reported pleasantness and the price of a wine, a result which invites an alternative perspective on the fundamentals of economic theorizing. This study investigates the importance of the price signal for this effect, relative to a non-monetary signal, by randomly assigning wine tasters to four different treatments. The wine used is from a wine category that does not fit immediately into most respondents' reference frame. Furthermore, the experiment supplements previous literature by adopting an incentive-compatible method. The results show a statistically significant but limited framing effect from the signals, including a complementary effect. In contrast with the standard view of incentives in experiments, the inclusion of incentives drastically leads to both higher valuations and a significant increase in variance, a result interpreted from a heuristic perspective. It is argued that this type of experiment serves to highlight the strengths and limitations of rational choice theory, helps to point towards a richer theory of economic behavior, and emphasizes the interactions between preferences, prices and expectations.

Acknowledgements

This work would not have been possible without the support of many. I am grateful to all friends and acquaintances who have helped me to carry out the experiments used for data collection, in particular all those who allowed me to borrow some of their time and resources for this purpose, including students and/or staff at the following institutions: Grythyttan Restaurangakademi, The National Defence Academy, The Military College Karlberg, and The Royal College of Music. Also, a big token of thanks goes to those wine clubs that invited me to their meetings.

I had fruitful discussions with my supervisor Jesper Roine at SSE/SITE, and also with Anna Dreber, SSE, about the experimental approach. James Tremewan (previously at SITE, now at the University of Vienna) provided crucial inspiration and guidance for the experimental setup and offered useful tips along the way. Furthermore, I am humbled and honored to have been given the chance to become part of the wonderful team at Pheasant's Tears winery that inspired this thesis in the first place. Finally, and most crucially, I am grateful for the invaluable support of my family, not least in facilitating the logistical aspects of carrying out this work. All inevitable remaining shortcomings are my own.

Disclaimer

The core of this thesis is an experiment involving wine tastings. The reader should know from the outset, that I am a co-owner of the vineyard that makes the wines used in the study. Since a purpose of the study is to measure perception of wine quality, there is obviously a risk of researcher bias in the results. I leave it to the reader to judge how this might impact the analysis.

A short background: In 2008, I was recruited by the Folke Bernadotte Academy, a Swedish government agency, to work as a civilian peace monitor in the European Union Monitoring Mission to Georgia. During my time in Georgia I became friends with an American living in Georgia since the 1990s, who had recently started a vineyard project with a Georgian friend. In 2010, they offered me to become a partner in the vineyard, and I became engaged in trying to reach out to other markets with the wine. In April of that year, I started a company in Sweden that is importing and distributing the wines to the European market.

Contents

List of Figures

List of Tables

1 Introduction

> Then to the Lip of this poor earthern Urn
> I lean'd, the Secret of Life to Learn:
> And Lip to Lip it murmur'd—"While you live,
> Drink!—for, once dead, you never shall return."
>
> – Rubaiyat of Omar Khayyam, v. 35 (E. FitzGerald, trans.)
> (Borrowed from the introduction to McGovern (2003))

> "This is a complex wine with qualities that maybe
> I don't know how to appreciate. For me as an ignorant,
> quality is that which I think that experts judge as quality."
>
> – Anonymous wine survey respondent

Humans have cultivated grapes with the purpose to produce wine since at least the dawn of agricultural civilization and its popularity shows no sign of abating. Wine is arguably one of the oldest traded goods still in circulation, and the industry is in steady development as a sophisticated wine culture spreads to new markets through globalization. For sure, wine has found its way into the economic and psychological literature as well, in such diverse areas as industrial cluster organization (Migone and Howlett, 2010) and the interplay between expectations and preferences, which is the focus here (Almenberg and Dreber, 2010; Lee et al., 2006; Veale and Quester, 2008).

As a consumer good, wine has some qualities that are intriguing from an economic perspective. It falls into a class of goods that are prone to conspicuous consumption, a term coined to describe consumption that is primarily aimed at signaling social status. Such goods have been called positional goods in the literature, following Hirsch (1977), or Veblen goods, in recognition of economist Thorstein Veblen, who first described the phenomenon (Veblen, 1899). Contrary to standard economic theory, the demand for a positional good can increase as it becomes more expensive, since consumption of the good serves the purpose of costly signaling of social status. Of course, quality goods can also be expected to command a higher price due to higher production costs, so that consumers should expect to associate price with quality, a result which has both a theoretical and empirical foundation (see section 2.1). Therefore it is useful to point out that the cost-induced correlation between price and quality is distinct from the signaling value of conspicuous consumption, although the two phenomena are naturally intertwined.

As will be expanded on in section 3.2, it has also been observed that prices in and of themselves not only affect perceptions of wine quality but also seem to elicit certain measurable

neural reward mechanisms in the human brain as the wine is consumed (Plassman et al., 2008). All this suggests a role for prices that goes beyond the classical market-clearing function in standard economic models, and opens up a venue for a class of interesting and rather pleasant experiments involving wine tastings; an excellent opportunity for the economist to combine work with pleasure.

1.1 Purpose and Method

The main purpose of this study is to make a contribution to the economic literature about wine and preferences by conducting an experiment where respondents in four treatment groups taste and evaluate a wine based on having received different information about the wine. The benchmark treatment group (group 1) received no information, while in the other three treatments respondents learned either the price of the wine (group 2), some facts, or conceptual information, about the wine (group 3), or both price and facts (group 4).[1] Against this setup, the key question to answer is whether the treatments have any meaningful effect on respondents' evaluation of the tasting experience. As alluded to in the title, it is also interesting to see whether the two signals complement or substitute each other. In contrast to some related studies, the experiment conducted here does not give any false information to participants.

The experimental design has a distinctly economic feature in that it incorporates a financial incentive into the evaluation process for all respondents except a control group. Another feature is that the wine used in the study represents a wine category that most respondents (and readers) are unlikely to have tried before, namely skin-macerated white wines that have been made in underground clay vessels in accordance traditions of ancient wine cultures, in particular from the Republic of Georgia in the southern Caucasus. Therefore, the taste does not easily fit into respondents' preexisting expectations. Finally the survey also attempts to measure some plausible alternative determinants for the perception of the wine, such as whether the participant has any formal training in evaluating wines.

1.2 Theoretical Framework and Outline

Although results from this type of experiment can be fascinating in themselves, the decision to place the analysis in an economic framework means that we should expect the results to tell us something about consumption and choice theory in general. Most standard economic models are based on a theory of rational choice, which should, in principle, be capable of predicting outcomes based on notions of scarcity and preferences, i.e. against a well-defined structure where resources are allocated based on optimality considerations. Strategic reasoning and social interactions can also be incorporated by

[1] In the following, italicized group references, like '*group4*', refer to the treatment *dummy variables*, while normal text, like 'group 4', refers to the treatment itself.

extending the analysis to a game-theoretical setting. The notion of equilibrium is crucial in such frameworks, and is best summarized as a state in which none of the economic agents in the model have any incentive to deviate from their current choice.

In real life, however, the incentives to deviate from a given choice can be countless. To start with, our tastes are not rigid, and can change rapidly or over the course of a lifetime, and for very complex reasons (Schelling, 1984). This author's sudden choice to become a coffee consumer in his late teens was in defiant violation of his personal taste, and almost exclusively a decision based on social considerations (although financial constraints certainly guided the choice since coffee was typically the cheapest item on café menus). It can be argued that the study of preference formation belongs squarely in the fields of psychology and the cognitive sciences and is only of secondary importance to economists. As a general rule this is probably the case, but economists should pay attention to such phenomena when preference formation is directly affected by economic variables and incentive considerations. Furthermore, everyday decision making among firms and individuals is typically clouded in uncertainty, restricted by time constraints, and influenced by habits and norms. It is not self-evident that an *a priori* assumption of optimality yields any deep insights about human behavior in such an environment. As will be argued here in the context of wine tasting experiments, the relevance of rational choice depends on the research question.

A theory that attempts to describe the type of behavior analyzed here probably needs to account for a very complex representation of the inner workings of the human mind, well beyond the scope of rational choice and game theory, which explicitly attempt to reduce the decision-making process to very simplified mathematical functions. The strength of these models lies in their simplicity, but they are probably misguided if the focus is to understand such phenomena as the formation of preferences, expectations and tastes, or heuristic decision making under uncertainty. Having made this point, however, it is maybe timely to reassure the reader that the results of the particular experiment conducted in this study do not represent any immediate challenge to the general validity of rational choice theory, although the study can arguably serve as an entry point to thinking about the boundaries of rationality. In that sense the study hopes to provoke some critical thinking about the resource-allocating mechanism of the price system and about choice theory in general.

The next section explores some theoretical aspects that are relevant for understanding the perception of wine. Section 3 complements the theoretical overview by reviewing some similar studies in the literature and relates them to the relevant theories. The method and experimental design are explained and commented on in section 4, which also includes description of the data and a short section about the Georgian wine tradition, which might be relevant for putting the study into context. Thereafter the results of the data analysis are presented in section 5, followed by a discussion and conclusions in sections 6 and 7, respectively.

2 Consumption and Choice Theory

Everyday experience tells us—at least the non-economists in the audience, if any—that prices, in and of themselves, affect preferences, expectations and behavior, a claim that will be elaborated in the context of wine consumption in section 3. In reality prices are more than just a mechanism in efficient markets. Prices convey stories. Arguably, the relationship between storytelling and prices can have a large impact at the aggregate level. Why would otherwise central bankers spend so much time on managing expectations and measuring *expected* inflation? Or why would luxury consumption be such a ubiquitous phenomenon across time and space? Indeed, going back to von Mises (1935), the very success of the free market in generating growth is often attributed to the improved coordination achieved by the rapid spread of information through market prices, so why not allow some room for a storytelling aspect of prices as well? Or, from a different perspective, acknowledge a role for storytelling as an alternative coordination mechanism for creating value? (Mitchell, 2005)

The social-status signaling associated with conspicuous consumption is only effective if the the value of the good used for signaling resonates equally with both the vain consumer and his intended audience. The good must possess a brand value with at least some general public recognition, or at least a market value that can be easily found. This is certainly true for some types of wines, such as Champagne, or Grand Crus from Bordeaux. The existence of a market value is actually crucial in achieving the signaling effect, unless the item's exclusivity is clear from the context, for example by being unique as a result of a famous provenance, like a piece of art. Suppose that a host wants to impress his guests by serving black Beluga caviar as an appetizer. Everyone knows that Beluga is expensive, which is enough to achieve the status signaling, regardless of its culinary value. However, the existence of a market value means that it would be redundant for our host to spend *more* than the going market rate for the caviar. Even if he did so, the action would not be directly observable by others, and even if it were, it would be unnecessary to pay more than the market rate since the signaling effect would be achieved regardless. Therefore, we wouldn't really expect the host to pay more than actually necessary, even for the conspicuous caviar. This argument serves to point out that even Veblen goods should be expected to abide by the law of demand, i.e. that even such goods are sensitive to price changes, although possibly at a much higher level than ordinary goods in a similar category. An extreme form of conspicuous consumption would be when money is spent visibly to others but in a wasteful way that destroys value. However, even in such cases the conspicuous, destructive consumer is presumably constrained by *some* budget, albeit an impressive one.

In any case, the typical consumer is probably acutely aware that wine also comes in all shades, from the preciously exclusive to the kind of mass-market product found on supermarket shelves all over the world. When a new wine is encountered, a consumer can be expected to try to categorize it somewhere along the quality continuum, and the price certainly informs that decision, but through intuition rather than rational

calculation, as will be argued below. What we have in wine then is a consumer good that has the potential to make the jump from an ordinary good to a Veblen good under certain circumstances. Understanding this mechanism is something of a holy grail for winemakers and producers in other categories of consumer goods, since it means that rather than competing on price in a competitive market, the winemaker can, entirely legitimately one might argue, extract monopoly rents from the brand value. A research question then naturally presents itself by trying to identify what it takes for the jump to Veblen status to occur.

2.1 The Strengths and Limitations of Rational Choice Theory

In the models of rational choice that have become staples of economic analysis, the concept of utility is understood as a more or less rigid structure in a person's worldview that helps her to filter potential payoffs so that the outcome is optimal for the given circumstances. Utility as such does not actually need to be given a positive definition. Rather, it is defined indirectly in these types of models as *that which is optimized* by the agent's choice. When a person chooses one option over an alternative, we take for granted that she preferred that option among the two, i.e. that there is a *preference relation* between the two options. The usefulness of utility as an analytical tool comes from the fact that, under certain technical conditions, such preference relations can be mathematically modeled as continuous real-valued functions (utility functions). The transformation of the binary preference relation into a continuous function might seem like an esoteric nicety for math geeks, but actually it represents a major contribution to the analytical toolbox: optimization problems, or choice problems, can be solved unambiguously when represented by certain types of real-valued functions. For example, such utility functions represent the consumers and firms in standard economic models, and is absolutely central to the notion of equilibrium in an economy. The ability to create and solve such models was arguably one of the great achievements of economics in the 20th century, and the formal justification for such models goes back at least to Debreu (1954).

A necessary, although possibly implicit, starting point for any analysis based on rational choice that attempts to explain a certain behavior is then an assumption about the validity of the axioms necessary to map preference relations into well-defined, real-valued functions. For many applications these assumptions seem natural and the resulting analysis yields reasonable results. However, for the purpose of this study it might be worth to dwell upon the meaning of the classical assumptions, since they become problematic in the experimental design employed here. So what are those assumptions? The purpose here is not to expand on this topic, but only to highlight some of the key characteristics of the axioms of consumer theory.[2]

[2]For a formal treatment, refer to an intermediate textbook on microeconomics; the presentation here draws on Jehle and Reny (2011).

The modern understanding of rational choice theory is concerned primarily with 'ordinal' utility as opposed to 'cardinal' utility. This distinction serves to point out that the theory does not assume that subjective preferences can be measured objectively, only that the *order* of preferences can be ranked. In contrast, the concept of cardinal utility lies closer to the hedonic understanding of utility as originally proposed by Bentham (1789). In modern theory, the basic building block is the notion of a 'consumption set', which encompasses all the possible consumption plans that the consumer can imagine. Furthermore, it is assumed that consumers possess the ability and knowledge to compare distinct alternatives ('completeness'), that the choice between alternatives is consistent ('transitivity'), and that preferences are not spontaneously reversed ('continuity'). It should be immediately clear that these assumptions can't be expected to hold in a situation where the consumer is subject to signals specifically designed to alter the imagination. Osborne (2009, p. 7) summarizes this argument:

> If we are considering how the markup of price over cost in an industry depends on the number of firms, for example, this sort of weakness in the theory may be unimportant. But if we are studying how advertising, designed specifically to influences peoples' preferences, affects consumers' choices, then the inadequacies of the model of rational choice may be crucial.

Another way to think of this limitation to the model is to consider what happens to our own worldview when we learn about some new gadget or invention. As a thought experiment it can be illuminating to consider, for example, how our own expectations about everything from everyday events to the feasibility of business models have been affected by the emergence of smart phones in the past decade. Most of us probably agree that this new technology has had a profound impact on our lives. The act of thinking and discovering is inalienable from our expectations and preferences. Our imagination is in constant flux.

Suppose for a moment, however, that preferences could be observed directly and incorporated into a model that would produce accurate predictions of a person's choice, for example in a wine buying decision similar to the experiment in this study. (As a practical matter this is precisely what corporations try to achieve when they collect data about consumers in order to fine tune future sales offers, so it is certainly a feasible concept.) Such a model would necessarily take the preference structure as given when evaluating alternatives. The present study is concerned with how choices are made when new information and concepts are introduced which do not immediately fit into the consumer's reference frame. A predictive model based on rational choice would struggle to predict outcomes under such circumstances. Not only would the model likely be inaccurate about outcomes, it would also fail to predict new preference structures. In this sense a model of rational choice is doubly unhelpful if our aim is to understand behavior when a person consumes a new, unfamiliar wine for the first time.

Rational choice is also a basic component of game theory, which in turn underpins many of the insights about signaling in economic theory. Game theory is successful in explaining a range of outcomes, especially referring to social behavior. As alluded to in the introduction, using price as a signal of quality can be understood as an equilibrium outcome in a signaling game when players have imperfect information. For a careful derivation of such a model, see for example Osborne (2009, pp. 336–340). In line with this theory, an empirical justification of the practical validity of price-quality heuristic is provided by Rao and Monroe (1989). Unfortunately for our purposes, however, such models have a weakness since they hinge upon the assumption that quality is well-defined and consumers know their tastes. A further limitation in the game-theoretic setup is that while signals can be a tool for coordinating on particular outcomes among many participants, they are typically not allowed to alter the original ranking of the possible outcomes in the models. On the upside, it is worth noting that game theory can help to illustrate how cultural norms, in line with an evolutionary view, can shape economic outcomes in a world where people, roughly speaking, belong to different personality categories (Ellingsen et al., 2012; Ostrom, 2000). Game theory can be very useful in understanding incentives and general behavioral traits and maybe even the evolution of norms and tastes, but it is not a precision tool for making accurate predictions about a specific consumption choice at a point in time. Nevertheless, section 3.3 uses a simple game to illustrate one possible way of thinking about wine tasting and valuation.

To expand on the point made in the previous paragraph, rational choice theory can also be extended to incorporate concepts like habit utility or adaptive expectations, which formally incorporate a role for preference formation over time (Englezos and Karatzas, 2009; Gorman, 1967; Pollak, 1970). As an example from applied work, Chow (1989) shows, using data on stock prices and dividends, that a central theoretical prediction from present value models, namely that current value is a linear function of conditional expectation of the next-period value, is consistent with an assumption of *adaptive* expectations, but not with rational expectations. On a theoretical note, it is even possible to include a role for Veblen goods in a general equilibrium framework, as shown by Eaton and Eswaran (2009). Extending rational choice theory to reflect such concepts surely yields valuable insight into how markets develop over time, but say little, if anything about the mechanisms of day-to-day decision making, such as judging a wine on the fly. Indeed, a motivation for such models in the first place is an assumption that decision making and preference formation are inherently complex phenomena. In short, as pointed out by Thaler (1985, p. 200), despite its successes, the economic theory of the consumer has not found any wide application in marketing (see however section 3.1 for an example related to the wine industry). There is a distinction to be made between understanding markets and understanding people.

2.2 Beyond Rationality: Heuristic Decision Making

If rational choice theory is not a suitable avenue for certain research questions, it is necessary to relax some of its assumptions and allow for a theory where outcomes are not necessarily viewed as the results of an individual's rationally weighing known alternatives against each other. Such an approach leads the economist to delve into domains that have traditionally belonged to psychology and the cognitive sciences. Needless to say, this opens up an entirely new field of research and the purpose here will only be to scratch the surface and offer a complementary perspective. Two of the earliest explorers of this approach were Daniel Kahneman and Amos Tversky. The former summarizes the benefits of this approach compared to rational choice theory (Kahneman, 2000, p. xvi, internal references excluded):

> A growing body of findings supports a radical challenge to the assumption, central to much economic theory, that stable preferences exist. The image of a decision maker who makes choices by consulting a preexisting preference order appears increasingly implausible. The alternative image is of a decision maker who chooses reluctantly and with difficulty and who constructs preferences in the context and in the format required by a particular situation. Of course, no one wishes to pursue the idea of context dependence to the point of nihilism. Choices are not nearly as coherent as the notion of a preference order would suggest, but they are also far from random. Some explanation of their limited coherence is therefore required. Perhaps people are better described as having attitudes than as having preferences.

This alternative view has been developed into a rich and coherent framework, sometimes labelled the "heuristics and biases literature," that offers many insights into human decision making both in market transactions and in other settings. Behavioral economics is a related term that has become popular within the economics profession, but for the particular context of wine tasting, 'heuristics' is a fitting label to capture this approach, and it will mainly be used here.[3] The theory allows for aspects of decision making that are intuitively compelling but incompatible with rational choice, such as time inconsistency of preferences, significant sensitivity to form and context (of particular relevance for this study) and decisions based on heuristic rather than rational reasoning. An overview of the field is offered by Kahneman and Tversky (2000).

The notion that context and presentation matter for individuals' decision making has been called framing, and is also a well-known concept in marketing. For example, Hoch and Ha (1986) showed that ambiguous *ex ante* evidence about the quality of a product increased individuals' susceptibility to advertising compared to when the evidence was

[3] Alternatives to rational choice theory for price formation have also been formulated in a sociological framework, where the individual is no longer central to the analysis, but where social and political forces shape market outcomes, for an overview see Beckert (2011). This approach is not explored further here.

unambiguous, while Levin and Gaeth (1988) noted that the framing effect tended to decrease when consumers were allowed to actually taste the product (beef). On the other hand, research on the *formation* of sensory expectations has been relatively sparse (Deliza and McFie, 1996; Lee et al., 2006). A perspective from cognitive science suggests a role for similarity, through inference-based reasoning, as an explanatory factor in human thought and perception, an idea that echoes the notion of habit utility mentioned in section 2.1 (Sloman and Rips, 1998). Studies involving various drinks, wine in particular, are of particular interest for the purpose here, and some papers that are especially relevant are reviewed further in section 3 (Almenberg and Dreber, 2010; Lee et al., 2006, and others).

Furthermore, results from this line of research have been used to distinguish between different categories of utility, in the economic sense, by introducing the concepts of 'decision' utility and 'experienced' utility. Decision utility refers to the understanding of utility in the modern economic sense outlined above, while experienced utility is a purely hedonic trait in the spirit of Bentham's usage of the term. Psychological research has shown that experienced utility can be characterized by a certain time inconsistency, so that retrospective evaluations are systematically different from immediate measures of current subjective experience, a result which has consequences for how we understand preference formation (Kahneman et al., 1997). Likewise, cognitive studies have shown that expectations influence actual subjective experience (Klaaren et al., 1994).

A recent trend that helps to structure this arguably very diverse literature has been to distinguish between two cognitive systems of human reasoning, one 'fast' system concerned with intuition and one 'slow' system focused on deliberation. Darlow and Sloman (2010) lay out both behavioral and neuroimaging evidence to support such a distinction in several fields of human reasoning, including categorization, formation of attitudes and moral judgement. The terms fast and slow were recently popularized by Kahneman, who also offers a very accessible and up-to-date overview of this developing field of research (2011). A summary of the heuristics literature is presented by Gilovich et al. (2002), which also explains the historical development of the field, going back to the initial publications by Kahneman and Tversky and other early contributors (Kahneman et al., 1982).

2.3 The Role of Incentives in Experiments

The role of incentives is a clear dividing line between economics and other social sciences. As economists we assume that incentives matter for individual behavior. This result is well-established in the literature and has become something of a paradigm for economic experiments. There is a lively debate about the role and purpose of incentives in experiments, and when to include them. The effect of incentives depends on the context, and in many settings they do not improve performance, but contribute to reducing variation in outcomes, especially for market-oriented contexts incorporating games, auctions and

risky choices, as documented by Camerer and Hogarth (1999), who provide a review of 74 studies involving incentives. According to their review, incentives typically improve performance if subjects in the experiment are required to perform effort-sensitive tasks such as solving puzzles, recalling items from memory or making predictions.

On the other hand, for some difficult tasks, incentives can actually reduce performance, typically if the task requires skills and knowledge that can not be acquired during the limited time span of the experiment. In view of the ambiguity of the effect of incentives, Camerer and Hogarth (1999) suggest that, in situations where results are known to be mixed, economic journals require that authors run multiple incentive conditions in order to publish the results. They also advance the view that experiments only measure short-run effects, "essentially holding capital fixed" (p. 35, capital should be understood in a broad sense, including cognitive capacity), but that more elaborate experimental designs could allow for long-run effects as well.

A separate review of incentives, with a focus on the provisioning of public goods, finds that economic incentives and social preferences can sometimes complement each other and sometimes work in opposite directions (Bowles and Polanía-Reyes, 2012). The authors highlight some mechanisms whereby incentives might influence preferences in a way that could potentially be relevant also for the present study, although wine tasting is distinct from public goods provisioning. Those mechanisms include: providing information about the person who implemented the incentive, framing the decision situation so as to suggest appropriate behavior, and affecting the process by which people learn new preferences. The general message is that incentives do matter, but it is not immediately clear from the vast literature on the topic exactly what effect we should expect in a wine tasting context. Thus concluding this general theoretical overview, the the next section narrows the focus to wine and beverage consumption in particular.

3 Understanding Wine Consumption

In this section some similar studies are briefly reviewed to put the present study into context. These papers offer different takes on the topic and differ in their theoretic emphasis along the lines outlined insection 2. The section also contains an attempt to complement the prevailing theoretical views with some considerations from a very simple game-theoretic framework.

3.1 Related Studies

It is straightforward to understand why people drink wine: it tastes good. What else to add? What other reasonable conclusion could we honestly hope for than to sip a glass of Chardonnay in the shade after the research is completed? There might not be any immediate answer other than that it combines work with pleasure. In any case, ample

research has indeed been carried out in this field, and there is even a Journal of Wine Economics to meet some of the demand.

A Conventional view of wine drinking that has produced a series of academic papers is to simply adopt a standard economic, rational-choice framework and attempt to tease out preference structures from observed consumer patterns, an approach that might be entirely legitimate if the purpose is to inform strategic decisions at a beverage business. For example, Veale and Quester (2008) conducted a wine tasting experiment where they attempted to measure the respective influence on quality perception from price, country of origin and taste. In order to achieve this they adopted an approach called conjoint analysis, a statistical method explicitly occupied with identifying preference structures that are assumed to exist, and which has been widely used by businesses (Green and Srinivasan, 1978, 1990). The method is quite interesting from a cost perspective in that it combines only a subset of the possible combinations of the variables in a way that leads to meaningful statistical conclusions without having to run every possible combination explicitly. The wines used in the study by Veale and Quester were all Chardonnays, but in different price categories and from different countries. In addition, the taste of the wine was manipulated slightly by adding a little bit of extra acidity in some cases. The general conclusion was that consumer belief about the quality associated with price and country of origin can trump the sensory experience. For example, a respondent was likely to give high ratings to an expensive French wine, even if the taste had been manipulated. Unfortunately for most winemakers however, relocating to France is probably not a feasible way of achieving Veblen status.

Lee et al. (2006) carried out an experiment where participants ranked a regular beer versus a special "MIT Brew" which contained a secret ingredient, balsamic vinegar—"a beer flavoring that most participants find conceptually offensive, but that does not, at this concentration, degrade the beer's flavor (in fact, it slightly improves it)" (p. 1055). There were three treatments: one group did not learn the contents of the secret ingredient, another group was told about it before tasting, and a third group received the information only after tasting, but before disclosing their preference. The result was, as expected, that disclosure of the secret ingredient reduced the preference for the MIT Brew, but, more importantly, that the timing of the disclosure mattered. Receiving the information before tasting the beer had a stronger negative impact on ratings, which suggests a role for framing in the perception of taste. This result directly inspired the treatment design of the present study. Lee et al. (2006) also review other studies with similar results, for example that alleged preferences for beers disappear when respondents can not see the labels of the beer they drink (Allison and Uhl, 1964). An overview of studies focusing on the interplay between sensory expectation and framing effects is offered by Deliza and McFie (1996).

Furthermore, Lee et al. (2006) use the terms top-down and bottom-up to distinguish between perception that is formed by external cues, like brand, or if it is generated by sensory experience, a theoretical dichotomy that fits well with the heuristic concept of humans as possessing different cognitive systems. A direct consequence of this theory

is that prior knowledge or information can affect the attention of someone evaluating a product. For this reason, the questionnaire in this study has been designed in an attempt to estimate how much time respondents spend tasting the wine.

An experiment similar to that of Lee et al. (2006) was conducted by Almenberg and Dreber (2010), who bring the concept of information timing into a wine tasting context. Corroborating the findings from the beer experiment, they find that timing matters for perception, but also note that the effect depends on the price level: if the wine is expensive, ratings are increased, while the opposite is not true. However, the effect only seems to be present among female respondents.

3.2 A Perspective from Neuroscience

As mentioned earlier, the heuristic or cognitive approach opens up a toolbox which has traditionally been far beyond the scope of economics. In particular, technology now makes it possible to scan people's brain activity while they concentrate on a particular task, and draw conclusions from the resulting patterns. Again, this perspective directs the analysis into a whole new area of research with countless applications, but this section will only highlight a few studies that are particularly relevant for a wine tasting context.

In order to try to investigate the mechanism that links marketing actions to individual decision making, Plassman et al. (2008) conducted an experiment where twenty participants tasted different wines while their brain activity was monitored by functional magnetic resonance imaging (fMRI). Subjects were told that they tasted five different Cabernet Sauvignons. They were asked to rate the taste of the wine, and the wines were labelled with a price tag that was used by respondents to identify the wine ($5, $10, $35 $45 and $90). Participants tasted the five wines multiple times and were asked to rate them after each tasting. However, the respondents did not know that the $5 and $45 wines, as well as the $10 and $90 wines were in fact identical.

It might strike the reader as a no-brainer (pun intended) that respondents would rate the $45 and $90 wines higher than their cheap counterparts, and this is indeed the observed pattern. One interpretation is that this effect is triggered by neural computations associated with the price-quality heuristic, i.e. that respondents are 'coded' to assume that expensive wines taste better. However, the results go further than that and show with strong statistical significance that the oxygen levels increase in a part of the brain which is associated with the encoding of *actual* experienced pleasantness. In other words, the neural representation of the tasting experience changes as a result of the stimuli combined with the price signal. Indeed it seems, after all, as though it is possible to measure hedonic utility objectively in this way. Furthermore, when the experiment was replicated without price tags, the effect disappeared. Similar effects have been observed when brand names are used as cues to respondents instead of prices. For example, McClure et al. (2004) found that both expressed preferences and measured brain responses increased

when people who tasted Coke or Pepsi received the drink in a brand-labelled cup compared to a control group who did not.

A more recent study by Salimpoor et al. (2013) used the same fMRI-based methodology to investigate brain activity when subjects were listening to music, but they also extended the analysis by incorporating an auction procedure whereby participants could buy music they liked. The music was not heard before by participants. The experimenters were able to establish that certain patterns of brain activity were positively correlated with how much the person was willing to pay for the music. Desirable new music elicited similar patterns as when listening to familiar, pleasant music. Furthermore, it has also been established with similar methodology that consumer choices are affected in a similar way both for a group of subjects that pay close attention to a particular product, and for a control group that were systematically distracted, thus suggesting that implicit, automatic processes might influence complex decisions (Tusche et al., 2010).

Note that in the wine study by Plassman et al. (2008) prices were communicated to participants by the experimenter, while the music study by Salimpoor et al. (2013) worked the other way around, so we are still faced with something of a chicken-and-egg puzzle when it comes to causality between price, value and experience. All in all, it seems as if though prices can be thought of as coordinating devices not only in market networks, but also in the neural networks of the human brain.

3.3 A Game of Expectations

In this section a simple game is considered with the purpose of complementing some of the results in the literature with a different view. The undeniable effect of framing on perception begs the question of whether the framing effect (like brand value or a price signal) is something that is valued on its own merit, or if it affects perception by modulating sensory-related chemistry in the body. If it is indeed a separate effect, then it could be understood in the context of social interaction, which would shed some light on the results in the neuroimaging literature presented above.

	Expensive	Cheap
Good	$t+s$	t
Bad	$t-s$	t

Figure 1: A Game Against Your Social Self

Consider the simple game in figure 1 to illustrate this point. (The labels Good-Bad and Expensive-Cheap are only meant as mnemonic devices to structure the argument, not as indicators of any intrinsic truth about the good consumed.) Think of a wine taster who does not know if the wine is expensive or cheap, but expects to find out the truth eventually. In the figure, we can think of t as the intrinsic sensory value experienced

by the consumer, while s is the social-esteem payoff from future social interaction. We can think of it as a two-stage game where in stage 1, t is revealed to the player, who then labels the wine 'good' or 'bad'. The second stage involves social interaction, and here the player must choose whether to reveal honestly her judgement from the first stage, in anticipation of the payoff s. In the first stage, then, the player expects the social interaction in stage 2, and is in fact making a judgement about her confidence to accurately identify the quality of the wine, and also about her ability to defend the choice openly. So the person's choice in stage 1 can depend on her expectations about her own behavior in stage 2. In fact, some wine tasting clubs use a similar approach when tasting wines: first everyone quietly tries a couple of wines and rates them according to his or her personal taste. After 15 minutes or so, participants are then asked to vote for the best and worst wine by raising hands. It seems obvious that peer pressure could influence tasting considerations in this setting. If you labelled the wine as good' and it turns out (through social interaction of some kind) that the wine you tried was indeed exclusive, then you gain social esteem through having correctly identified the sublime nuances. On the contrary, labeling an 'expensive' wine as 'bad' could put you in an awkward situation by revealing your ignorance. Labeling a 'cheap' wine as 'good', however, is not as risky for your social esteem, since that can be regarded as a kind of default option that should not be controversial (in fact, "value for money" seems to be the dominating paradigm in mainstream-media wine reviews). This argument is also motivated by the results of Almenberg and Dreber (2010), who document precisely such a pattern. On the other hand, there is also no reward from labeling a cheap wine as 'good.' The game captures the idea that all the risk to the ignorant wine taster is on the downside, and that playing 'good' can always be a safe strategy. Note that if a price tag is used to identify the tasted wine, as in Plassman et al. (2008), then the strategy of playing 'good' whenever the wine is expensive would (trivially) be the only Nash equilibrium in this setting, since the player would have no incentive to deviate, assuming that the information is true. It is a theoretical possibility, albeit speculative, that unconscious equilibrium considerations of this kind could generate certain cognitive patterns that might show up as positive correlations in neuroimaging studies.

Figure 2 (page 15) presents a variation of the game where the player's social circle consists of wine snobs and where more social esteem is derived from wine bragging than among the average wine consumers of figure 1. Now there is a social esteem risk to labeling a 'cheap' wine as 'good', but also a reward to revealing a cheap wines as 'bad'. Furthermore, the wine snob friends derive a *schadenfreude* utility of f from witnessing the faltering judgement of a fellow wine snob. If the price tag is set by a neutral third party (the retailer or the experimenter), the wine snob friends only observe the outcome without participating, and the game now has two equilibria: play 'good' if the price tag is 'expensive' and play 'bad' if the wine tag is 'cheap'.

If, however, the wine snob friends are responsible for setting the price tag, they participate in the game, and the outcome will depend on the players' believes about each other. This could be explored formally as an extensive game with imperfect information in the spirit

	Expensive	Cheap
Good	$t+s, 0$	$t-s, f$
Bad	$t-s, f$	$t+s, 0$

Figure 2: A Game of Wine Snobs

of Osborne (2009) as mentioned in section 2.1 above, but for the present study, the simple games presented here serve to emphasize that explicit or implicit social interactions can potentially affect a tasting experience. This framework also captures another idea that will be tested in the data analysis, namely that those with special knowledge about wine can expect to be more confident in judging the wine. Their status as wine connoisseurs implies that they can possibly influence others and modulate the value of s through their own behavior. Of course, a wine professional could be expected to identify wines more accurately than the average person, but here the point is to highlight a possible effect from social interaction.

A relevant question then is whether the correlation between price and experienced pleasantness identified by Plassman et al. (2008) is a symptom of large t or large s. Unfortunately we have no straightforward way to test the hypothesis H_0: $s = 0$. Such a test would be important for methodological purposes, since a rejection of H_0 would warrant a game theoretic explanation of the observed behavior, while we would have to resort exclusively to psychological or cognitive theories if we fail to reject it. Without resorting to any pre-drawn conclusions, it is worth noting that wine consumption is by its very nature is a social phenomenon with strong cultural connotations.

Note that this simple game with one's self in anticipation of having to defend the choice against prevailing norms in a series of future social interactions, could potentially capture both the pleasing effect of price reported by Plassman et al. (2008) and the reward effect of music reported by Salimpoor et al. (2013). If so, the player might be searching, consciously or not, for both extrinsic and intrinsic cues in the consumption experience that might inform her about the context of the good. Now, this idea might be far too simplistic to explain a phenomenon guided by the immensely complex neural mechanisms related to sensory experience, but serves to highlight the difference between two tentative explanations of what is going on: according to one perspective, the signal works to trigger certain internal cognitive mechanisms in the brain that would otherwise be dormant, thereby modulating the actual sensory perception. Alternatively, the signal is seen as a trigger of certain social considerations that might, in turn, inform a heuristic judgement process in the spirit of Kahneman and Tversky (2000). Of course, these two views are not mutually exclusive, since both measures, t and s, can be modulated independently, but by combining a game-theoretic approach with neuroimaging, it might be possible to design an experiment with refutable hypotheses about the measures.

4 Experimental Design and Theoretical Predictions

This work was initially spurred out of curiosity over an observation made while organizing wine tastings for a commercial purpose with the wine used in this study. There seemed to be a tendency that an individual's perception of the wine was influenced by his or her general knowledge about wine. More knowledgable persons seemed to appreciate the wine more than others. Similarly, perceptions seemed to be more favorable when respondents had a chance to first get an introduction to the wine. As an initial experimental idea, this relationship could have been investigated by looking at how knowledge and information affect preferences and perceptions. Concretely, the purpose would have been to look for any significant difference in perceptions depending on respondents' formal training in tasting wine, or receiving information about the wine. The present study is the result of a rather lengthy process of discussing this phenomenon with others, including economists, and reviewing some of the literature referred to above. The initial design idea was quickly rejected, however, since it would be difficult logistically to organize a study where the experimental design presupposed having control groups where respondents were selected based on their knowledge. Instead, the focus shifted to the general effect of the signal treatments (monetary versus non-monetary signal), and also came to incorporate incentives.[4] The key variable used to capture the effect is the respondents' valuation of the wine, labelled *ownvaluation*.

4.1 Treatment groups

As mentioned in the introduction, there are four signal treatments, and to apply them experimentally, four different questionnaires were used, each corresponding to one treatment: (1) a benchmark treatment where respondents are given no extrinsic cues about the wine prior to tasting, (2) a price treatment where respondents are informed about the retail price of the wine prior to tasting, (3) a story treatment in which respondents are exposed to an informative text about the wine prior to tasting, and (4) a combined treatment which combines the information in treatments 2 and 3. The text used in setting (3) is based on facts, but has been purposefully written to be interesting for the participant. The story treatment should be thought of as a conceptual signal, and, needless to say, it could have been modified in a myriad different ways. The main purpose of this treatment is *not* to draw conclusions about that story's particular relevance, which would be nonsensical, but rather to give the respondent a legitimate rationale for why the wine they taste is so (presumably) different from what they have tasted before. The design of the questionnaire is presented in appendix B on page 52, together with some additional comments about the logistical aspects of carrying out the experiment.

At each experiment session questionnaires were assigned randomly to respondents. All questionnaires were equipped with a cover sheet that was identical for all settings, so

[4] I am especially indebted to James Tremewan for pushing me in this direction and giving helpful advice about how to elicit incentive compatibility in practice.

that the experimenter wouldn't know in advance how questionnaires were distributed. Respondents were explicitly encouraged not to communicate with each other during the tasting. This guideline was not strictly followed, but arguably all sessions came reasonably close. A more careful experimental design would have ensured that participants were isolated from each other during the tasting, but at the cost of sacrificing realism. As it turned out, the study can be viewed as incorporating aspects of both lab and field experiments. For a discussion of the distinction between the two, see for example Harrison and List (2004), who argue that the careful experimenter should be aware of differences in outcomes resulting from the experimental design. However, no attempt has been made to categorize observations along that distinction in this study.

Randomization of treatments serves the purpose of averaging out any other possible aspects that might influence results. This paradigm is a cornerstone of credible statistical analysis in general, and in the social sciences in particular, where the potential for omitted variable bias is unlimited, especially when there is no explicit theoretical model against which to scrutinize the data, as in the natural sciences. Therefore, the present study comes reasonably close to the experimental ideal when it comes to the signal treatments. Caution is however warranted for the incentive treatment, since it is not randomized at the individual level, but across sessions, which should be taken into account in the data analysis. For a discussion about the experimental ideal, see Angrist and Pischke (2009), especially chapter 2.

However, randomization is not a panacea for understanding the world, although it can help to tease out causal relationships in otherwise complex circumstances. As pointed out by Deaton (2009), random trials are often context-dependent, and defer attention away from actually understanding the underlying mechanism of *why* something works, rather than simply establishing *that* it works. Deaton also criticizes the domination of randomization in applied work for its tendency among researchers to focus only on research questions that are suited to a particular experimental design. With this in mind, we should be careful not to draw any strong conclusions about statistical correlations unless firmly rooted in theory.

As for the actual experiment, in total 289 people participated at around a dozen different sessions at various places in Sweden in the spring of 2012 and in the period between December, 2012 and May, 2013. There were between 67 and 82 respondents in each of the four signal treatment groups. Of the total, 42 participants were subject to the non-incentivized version and the rest to the incentivized treatment. However, after an initial incentivized session, the design of the incentive elicitation mechanism had to be changed, because it turned out that up to 50% of respondents failed to understand the key question about how to value the wine. This result is interesting in itself and is explained further in section 4.3. In total, 213 respondents completed correct questionnaires (with 2 observations for *ownvaluation* missing for unrelated reasons) with the incentive treatment, of which 49 came from the initial design, and 164 from the revised design. 34 observations on *ownvaluation* are treated as missing, see section 4.4 for more details.

4.2 The Georgian Wine Tradition

The republic of Georgia in the south Caucasus hosts the world's oldest archeological findings of winemaking from cultivated grapes, and it has recently been scientifically established that this region is the likely origin of the domestication of the wild grapevine into its cultivated cousin, the *vitis vinifera* (Johnson, 2004, pp. 13–14). For a fascinating, scholarly account of the ancient origins of winemaking, see McGovern (2003). A traditional Georgian white wine is so different from a classical white wine that it cannot be considered a close substitute for such a wine, as elaborated in the questionnaire shown to wine tasters in groups 3 and 4 (page 52). The key aspect of this difference is that, unlike a conventional white wine, in the Georgian tradition the grape skins are left with the grape juice during and after fermentation. Thereby, a Georgian white wine might be thought of as sharing more similarities with a conventional *red* wine, which gets its color precisely from such treatment. Typically very few people make this connection while tasting the wine, but a majority immediately recognize the similarity to red wines when it is explained to them. The wine used in the study was a 2009 vintage of the grape 'Rkatsiteli' from winemaker Pheasasnt's Tears in the Georgian wine district of Kakheti. The retail price of this wine was 139 kronor (a little over $21) during 2012, but was increased to 150 kronor by the end of that year. Potential effects of the price hike for the purpose of the study are discussed in section 4.5. As for the "actual" quality of the wine in terms of reviews by professional wine critics, this wine, or similar wines from the same producer, have been judged favorably, and are served at some top-notch restaurants in places like London, Paris and San Fransisco (Robinson, 2012; RVF, 2011).

4.3 Elicitation of Incentive-Compatible Valuations

In order to induce respondents to give a reply that is consistent with their actual preferences, rigid or not, it is necessary to design the experiment accordingly. The explanation that follows here mirrors closely what was actually communicated to respondents verbally before the tasting and in written instructions on the questionnaire (p. 54). One question simply asks the respondent "How much would you be willing to pay for a bottle of this wine?" Respondents marked their reply on a scale divided into multiples of five from 40 to 200 (or 50 to 210, as explained in the next section). They were also told that they had a chance to win a prize after the tasting in the form of either a bottle of the wine they had tasted, or a sum of money. The winner was drawn in a lottery among all participants at each session. Questionnaires were numbered in order to administer the lottery. Furthermore, respondents were informed that the type of prize (wine or money) depended on their answer on this question in the following way: once a winner was drawn, the experimenter would randomly draw one of the 33 numbers on the scale from 40 to 200 (or 50 to 210), and compare it with the winner's answer on that question. If the drawn number was lower than or equal to the winner's answer, the prize would be wine, but a higher number meant that the winner got the corresponding sum of money. In that way the winning respondent would be guaranteed to get the wine bottle if the

valuation was set at the maximum boundary of the scale, thereby indicating a valuation of (at least) 200 kronor (or 210 kronor in some sessions).

To check for an incentive effect we obviously need a control group, so some respondents were not given the incentivized version of the questionnaire. A note of caution is warranted. Unlike the signal treatments which were randomized at the individual level, the incentive treatment was not independent within sessions. By the nature of how the incentive elicitation was designed, all respondents at a given session were either incentivized or not. A way to deal with this challenge is to used clustered standard errors in the regression analysis.

As mentioned above, up to 50% of respondents clearly had trouble understanding the valuation question in the first design. In the first design (which is not reported in the questionnaire in the appendix), respondents were confronted with 33 choices of the form: "would you prefer to win 40 kronor or the bottle?", "... 50 kronor or the bottle?", and so on up to 200. This approach should in principle lead to the same answers as the simpler method described above. However, the complicated question design, coupled with the nested probabilities in the lottery ("*if* you win, *then* a random number is drawn..."), caused some respondents simply ignore the question and leave it empty. Some respondents only answered one of the 33 questions; a forgiving interpretation would be that they simply marked the cut-off value where they valued the wine, but in order to avoid any pre-drawn conclusions, such answers are treated as missing observations. More intriguingly, a non-trivial fraction of respondents (11%) gave answers that seemed to be *random*, not unlike the betting pattern you would observe on a roulette board. For example, someone could mark that they preferred 40 kronor rather than the wine, but the wine rather than 100 kronor, and so forth. It might be added that the pattern was observed in all treatment groups *except* the benchmark treatment. Apart of frustrating the experimental economist, this clearly indicates that respondents in general had trouble categorizing the wine. It also serves to bring home the argument of Deaton (2009) about researchers focusing narrowly only on research questions that lend themselves to certain methods. Rather than molding the data to fit the researcher's purpose, a more honest approach might have been to change the research question. Although the problem of non-compatible answers did not occur in the second design, in view of this complication it is necessary to accept the likely possibility that some respondents, even in the revised design, either did not understand the information given to them, or did not read the information carefully, which would obviously be a problem since the treatment effects require that the respondents read the material in the questionnaire. The overall impact of this ambiguity would be to reduce the treatment effects.

4.4 Data Description

All data in the analysis comes from the survey used at the wine tasting sessions. To understand the data gathering process, it is recommended to have a look at the questionnaire in appendix B. The primary dependent variable in the analysis is *ownvaluation*,

which was introduced in the section about incentives above. In order to allow for a richer analysis, data was collected on respondents' judgement about how others value the same wine. The valuation variables are also complemented by questions about quality, taste and even inclination to buy a bottle of the wine in future, all of which are intended as proxies for valuation. An extensive presentation of the summary statistics for most variables is found in table 13 in the end of appendix A on page 50.

Here it is timely to mention that a little flaw in the experimental design introduces a potential bias. Respondents at incentivized sessions marked their reply by ticking a box on a scale from 40 to 200 (or 50 to 210, as will become clear in section 4.5), but at non-incentivized sessions they were simply asked to write down a number. The decision to limit the valuation variable to a rigid scale was motivated by the method used to elicit incentive compatibility, as explained in section 4.3. Ideally, the non-incentivized sessions would have used a similar scale for maximum comparative power. The reason why this was not implemented was that the initial sessions were non-incentivized, and the final design of the incentivized experiment grew out of experience from those first sessions. Once the distinction was made, it was decided to stick to the design and deal with the implications at the analytical stage, but if the experiment were to be repeated, this could easily be avoided. A straightforward way to deal with the issue is to assign the minimum value from the scale (40 or 50 depending on the session date) as a replacement to *ownvaluation* for all observations with *ownvaluation* < 40 (or < 50). This new variable is labelled *ownvaluation2*.

The main treatment variables are called *group2*, *group3* and *group4*, but variations are also considered where the signals (*price*, *story*, and *bothsig* for 'both signals'), rather than the treatments, are used as dummies, so that *group4* would be replaced by $price * story$, as explained further in the results section. The variable *bothsig* has perfect collinearity with *benchmark*, and the only effect of exchanging one for the other is to reverse the sign on the coefficient. For ease of exposition, *bothsig* is used in the regression tables with the results (section 5), since it ensures that all treatment variables have positive signs.

It seems reasonable that other variables than the ones of primary concern could have an effect on perception and valuation outcomes. One hypothesis based on casual observations prior to carrying out the experiment suggests a potential negative effect from age. Likewise, it seems likely based on experience that having any formal education related to wine could have a positive effect, and some information of this type is also collected through the questionnaire.

4.5 Checking for an Inflation Effect

Since the actual price of the wine (*actualpr* in table 13, appendix A) increased from 139 to 150 kronor during the course of the experiment, it might be relevant to ask whether the average respondent's self-reported typical spending on wine increased during this period,

i.e. if there was an inflation effect.[5] If, for brevity, we denote the respondents' usual spending *usualprice* by p_u and the actual retail price *actualpr* by P_a, we can formulate the null hypothesis of no inflation effect as:

$$\mathrm{H}_0 : \delta = 0, \text{where}$$
$$\delta = \mathrm{E}[p_u | P_a = 139] - \mathrm{E}[p_u | P_a = 150].$$

Using the mean as an estimator for the expected value E, a standard, two-group mean-comparison t-test can be used to test H_0. However, in order to perform the right test we must make an assumption about whether the variances are equal in the two groups. To check this, let σ^2_{139} and σ^2_{150} denote the respective variances. There are 176 observations in the first group and 91 in the second. If the true variances are indeed equal (which we cannot know for sure), it would be true that $\sigma_{139}/\sigma_{150} \sim F(175, 90)$, a distribution which has its 99^{th} percentile at 1.557. For our sample, the F-statistic is calculated to $1.807 > 1.557$, so we reject the null of equal variance at the 1%-level. However, a quick look at the data shows that there are two outliers in the first group with $p_u = 300$. If we drop those observations from the test, the variance ratio now follows an $F(173, 90)$ distribution.[6] The new F-statistic is 1.0324, which is within the distribution's 57^{th} percentile, clearly suggesting that the variances are actually equal. With this in mind, we perform a t-test with equal variances and fail to reject H_0 against a two-sided alternative ($t = 0.45$, or $t = -0.17$ if the outliers are excluded), so there doesn't appear to be any detectable inflation effect in the data. It is reasonable to conclude that the only impact of the change in the actual retail price is to make the price signal stronger, which only serves the purpose of the study.

A second complication from the increased price concerns the design of the questionnaire used in the survey. Remember that the price is included as a piece of information for participants in treatments 2 and 4. As already explained, the task of administering incentives to respondents required that they mark their valuation on a truncated scale divided into multiples of five. In practice this meant that respondents marked their choice by ticking one of 33 boxes. Since the scale is bounded at a minimum and maximum value, there is a chance that the location of the boundaries influences the respondent's choice, which in turn risks blurring the effect of the price signal that we are interested in. In order to compensate for such an effect in the questionnaires used in later experiments, the boundaries of the scale were increased by a similar amount as the retail price, i.e. by 150 - 139 = 11 ≈ 10. When the actual price was 139 kr, the scale ran from 40 to 200, and from 50 to 210 when the price was 150 kr.

[5] The increase in the retail price of the wine in December, 2012, was in fact Your experimenter's personal decision, and was motivated purely by business considerations related to distribution costs.

[6] The outliers came from data gathered at a session with sommelier students; no session with sommeliers was carried out after the price increase.

We can further guard against potential problems arising from the price increase by performing some robustness checks in the data analysis, such as using an alternative dependent variable, *adjval*, where respondents' valuations have been reduced by the minimum value of the scale (for exampe, a valuation of 110 would be replaced with $110 - 40 = 70$ when the actual price was 139, and with $110 - 50 = 60$ when the actual price was 150). This causes problems with interpretation however, especially since some values will become negative. It might also upset a potential incentive effect on valuations. Another approach to tackle this issue is to also run control regressions only for subsets of the data with $actualpr = 139$ and $actualpr = 150$.

4.6 Theoretical Predictions

To reiterate the main purpose, the experiment in this study asks participants to value a wine after they have tasted it, while subjecting participants to different treatments (two signal treatments and an incentive treatment). It should be clear from the context that the treatments are expected to have a positive effect, if any. This is a reasonable hypothesis based on previous studies and the theoretical considerations presented earlier. This assumption is relevant for hypothesis testing in the results section, since it means we should focus on one-sided alternative hypotheses, of the form $\mathrm{H_a} : \beta > 0$, for the signal treatment variables.

Rational choice theory suggests that tastes and preferences are rigid, so we should expect little effect from the treatments since respondents are allowed to taste the wine before forming an opinion. However, since economic agents are assumed to know their tastes under this theory, the variation in respondents' valuations of the wine they taste should not differ significantly from the variation in what respondents typically pay for a bottle of wine. After all they are allowed to taste the wine and simply have to consult their preferences. Under rational choice, it seems reasonable to assume that variation of wine expenditure in general for a given population sample should not differ significantly from the variation in valuations of a particular wine.

The heuristic or behavioral approach on the other hand suggests a role for treatment effects, in particular through framing mechanisms. However, the finding of Levin and Gaeth (1988) referred to above suggests that framing effects decrease when respondents are allowed to taste the product, so this effect might work in the other direction. But suppose there is a treatment effect from the monetary and non-monetary signal, what patterns should we expect to observe? If the signals are complements, we should see statistical significance for all treatment groups that received any signal. In particular, the treatment for group 4 (subject to both signals) would be positive and significant even when controlling for the effect of groups 2 (price only) and 3 (story only), thus indicating that the signals reinforce each other. If, on the other hand, the signals substitute each other, then group 2 and 3 would show statistical significance while group 4 should be insignificant. Furthermore, the fact that the wine used in the study is likely to be

unfamiliar for most respondents should lead to significant variability in perception from a heuristic perspective, reflecting judgement under uncertainty, or perhaps indicating that the lack of familiarity divides respondents into consulting different cognitive processes to evaluate the new sensation.

As for the incentive treatment, a natural hypothesis would be that incentives reduce variation in results, which would be in line with well-established results in the literature, mainly through the coordination mechanism of the price system which most people implicitly have internalized by simply living in a market economy. It is not clear that incentives should affect the sample mean valuation of a wine in any particular direction, so we should focus on two-sided alternatives for this treatment variable. However, any effect on mean valuation would suggest that incentives in themselves trigger certain cognitive processes that are otherwise dormant. Furthermore, it might be expected that incentives interact with the price signal to strengthen the coordination mechanism of prices, so it might be relevant to look for interaction terms between incentives and the other treatments.

Finally, neuroscience offers the exciting possibility of actually trying to predict outcomes for individual respondents in the spirit of Salimpoor et al. (2013) and Tusche et al. (2010), but this hypothesis unfortunately remains illusive for this experiment's necessarily limited budget.

5 Results

This section introduces the main findings of the experiment by presenting some summary statistics and a series of regression tables. Before investigating the size of the treatment effects, it is worth mentioning that the result which stands out most clearly in the statistical analysis is not captured by significance levels and coefficient values; namely, that the variation in respondent valuation of the wine is large across all treatments groups, and is significantly larger than the variation in what respondents typically pay for a bottle of wine, indicating that respondents differ greatly in their perception of the wine.

Table 1: Variance Comparison

	Obs.	Mean	Std. Dev.	95% Conf. Int.
ownvaluation	253	93.77	38.33	89.03–98.52
usualprice	267	95.77	30.52	92.09–99.45
othervaluation	284	91.46	30.33	87.92–95.01

Table 1 compares the mean and standard deviation for three variables related to the valuation of the wine. The first variable is the respondent's valuation of the wine,

ownvaluation, and the second is the amount of money respondents report to typically pay for a bottle of wine, *usualprice*. The similarity of the means of those two variables suggests that maybe *usualprice* is an important determinant of *ownvaluation*, so it will be included in the control regressions. On the other hand, the standard deviations are strikingly different. A variance ratio test can be used to check the null hypothesis of equal variances. This test gives an F-statistic of 1.5776 while under the null hypothesis the ratio follows an $F(252, 266)$ distribution with its 99^{th} percentile at 1.3357, so we firmly reject the null hypothesis of equal variance. The third variable in the table is the respondents' assessment of how much other people value the wine on average, *othervaluation*. It is clear from table 1 that *ownvaluation* has a different distribution than the other two variables. To illustrate the difference, histograms of those three variables are presented in figure 3, together with the variable $d\_usual = ownvaluation - usualprice$.

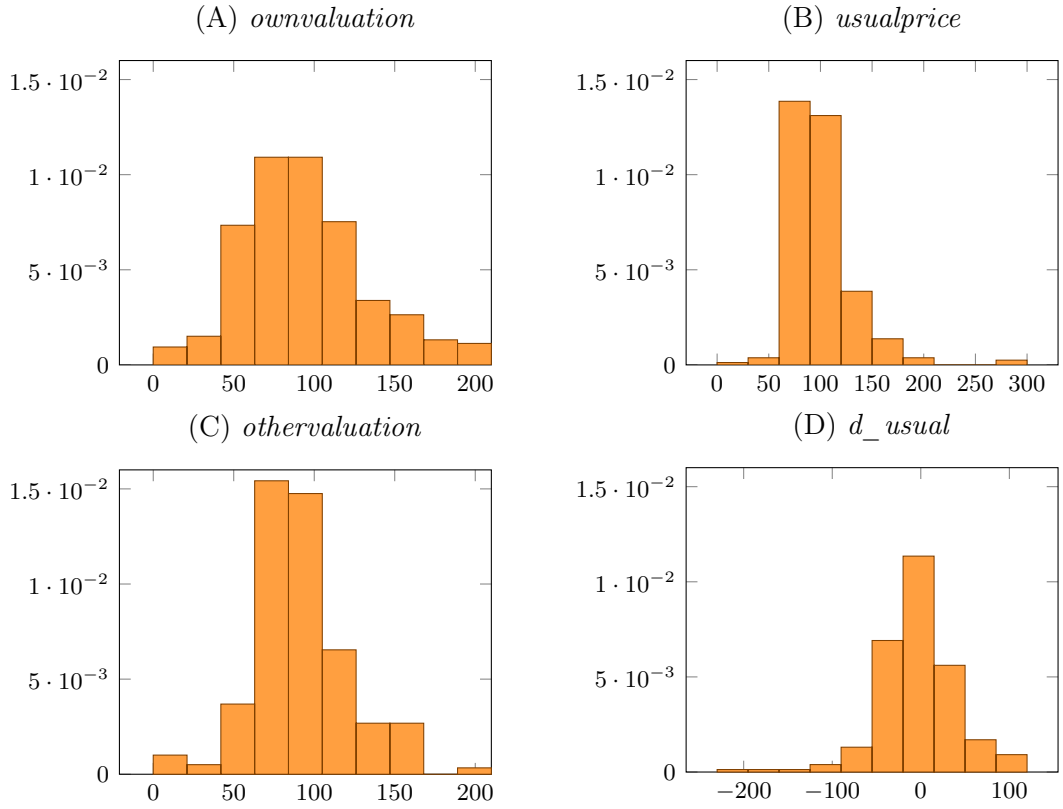

Figure 3: Some Distributions of Valuation Variables

The x-axis shows the variable values, and the y-axis shows the distribution frequency (total area sums to unity for each histogram). Note that the scale of the x-axis is not identical in the four diagrams (and hence bar widths are also different), so comparisons should focus on the frequency denoted along the y-axis. The figure shows that the distribution of *ownvaluation* (A) has thicker tails than the distributions for *usualprice* (B) and *othervaluation* (C). Note also that the only effect of switching from *ownvaluation* to

ownvaluation2 is to bundle the tiny, leftmost bar of histogram (A) into the next two bars of (A), so it does not affect the general shape of the distribution much (not included).

Table 2: Mann-Whitney Test for Treatment Effects

	Tested variable: *ownvaluation*2				
No. obs.	73	55	55	70	180
Mean	87.93	95.05	100.27	96.63	97.26
MW p-values	*benchmark*	*group2*	*group3*	*group4*	*bothsig*
benchmark	—				**0.0685**
group2	0.2769	—			—
group3	**0.0986**	0.5817	—		—
group4	0.1133	0.7199	0.8753	—	—

The table reports two-sided Mann-Whitney p-values for testing the null hypothesis that *ownvaluation2* is different between treatments. Values below 10% are marked with boldface. One-sided p-values are obtained by dividing the table values by two (relevant for 1st column and 1st row).

Having established that variances are very pronounced across treatments, let's turn the attention to the size and statistical significance of the various treatments. First the main results are summarized in two tables which look at mean values of the variable *ownvaluation2* within the different signal treatments groups (table 2) and between the incentivized and non-incentivized subsets of the sample (table 3, page 26). These two tables also present the two-sided significance levels from a standard Mann-Whitney test comparing the treatment effects against each other. These tables help to summarize the results which will be explored further in a regression framework in the remainder of this section. The advantage of the Mann-Whitney test compared to regression analysis is that it does not impose any structure on the sample data, but the drawback is that it does not capture the importance of the variation in results.

As can be seen in table 2, the mean valuation among respondents who received the conceptual signal (treatment group 3) is significantly different from the benchmark group. Likewise, respondents who received either one or both signals (captured by the dummy variable *bothsig*) also exhibit a statistically different mean valuation. Allowing for one-sided p-values in comparisons against the benchmark group (relevant for column 1 and row 1) means that the significance levels should be divided by two, which renders treatment group 4 statistically significant at the 10%-level, and group 2 at the 14%-level. This indicates the presence of treatment effects, including a complementary effect captured by group 4.

Furthermore, table 3 shows that the the incentive treatment has a large and rather unexpected effect, both statistically and in magnitude. The incentive treatment is more pronounced than the actual signal treatments in the previous table. Overall, this result

Table 3: Mann-Whitney Test for Incentive Effect

Variable: *ownvaluation2*	Total	Mean within Treatment Group *benchm.*	*group2*	*group3*	*group4*	*price*	*story*	*bothsig*
Incentivized	98.18 (211)	91.07 (61)	96.17 (47)	109.64 (42)	98.93 (61)	97.73 (108)	103.30 (103)	101.07 (150)
Non-Incentivized	76.45 (42)	72.00 (12)	88.50 (8)	70.00 (13)	81.00 (9)	84.53 (17)	74.50 (22)	78.23 (30)
MW p-value	0.0004 ***	0.047 **	0.650	0.002 ***	0.219	0.219	0.001 ***	0.092

$*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$ (Two-sided). No. of obs. in parentheses.

spurs a curiosity about whether the different treatments have contributed to the increased variance in *ownvaluation*. Somewhat surprisingly, the next section shows that regression models for the signal treatments exhibit only limited explanatory power (low R^2), even when several controls are added, including *usualprice*. Similarly, it will be shown that the incentive treatment also explains only a fraction of the overall variation illustrated in figure 3.

5.1 Baseline Regressions

This section presents two tables with baseline regressions that complement the Mann-Whitney tests in the tables above. To start with, table 4 on page 27 shows four basic regressions, with *ownvaluation* as the dependent variable. This is the baseline result and will be used for comparison with other specifications. Two things are worth noting about how the regression tables are organized. Firstly, the four columns represent four different model specifications, with slight variations. The first column contains the variables *group2*, *group3* and *group4*, which refer to the various treatment groups. As explained above, respondents in group 2 received the price signal, respondents in group 3 received the non-monetary signal, and respondents in group 4 received both signals. Group 1, which is excluded due to perfect collinearity, is the benchmark treatment group that received no signal. In order to understand what's going on in the data, keep in mind that the variables in column (2)–(4) are only linear combinations of column (1), and are included for comparative purposes. Essentially, in column (3), *group4* has been absorbed into *price* and *story*, while in column (4) all treatments are grouped into one single dummy variable. A second thing to note is the inclusion of some summary statistics at the bottom of the table. The F-statistic refers to the joint significance of all variables in the respective specifications. N is the number of observations in the sample.

A potential interaction effect between the monetary and the non-monetary signal is

of particular interest, and, as already mentioned, it is captured by the treatment assigned to respondents in group 4. An alternative approach would have been to look at an interaction term of the form $price * story$, but this term is excluded from table 4. To understand why, note that $group4 = price * story$ but that, by construction, $group4 \neq group2 * group3 = 0$. This fact is illustrated in table 8 in appendix A (page 45), where the first two columns show that those two specifications are identical. The third column in that table illustrates that *group4* is absorbed into *price* and *story*, which causes the coefficient on the interaction term to become negative. Thus, the interaction term $price * story$ is redundant in the analysis. Note also how the constant term is identical in all four columns of table 8, unlike in table 4.

Table 4: Baseline Regressions

Regressand: *ownvaluation*	(1)	(2)	(3)	(4)
group2	7.927			
	(1.16)			
group3	12.072**	6.909		
	(1.77)	(1.18)		
group4	8.376*			
	(1.31)			
price			2.168	
			(0.45)	
story			6.314*	
			(1.30)	
bothsig				9.368**
				(1.77)
_constant	87.110	92.273	89.584	87.110
R^2	0.014	0.006	0.008	0.012
F-stat	1.169	1.401	1.036	3.128
Prob $> F$	0.322	0.238	0.356	0.078
N	253	253	253	253

$*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$ (One-sided)
t-values in parentheses.

The primary insight from table 4 is that the treatments only explain a tiny fraction of the overall variation in *ownvaluation*, as indicated by the minuscule R^2 values. In the first specification, *group3* is statistically significant at the 5%-level, and *group4* at the 10%-level. The coefficient for *group2* is not significant at the 10%-level, but the t-statistic of 1.16 corresponds to a p-value of 13%, which, at best, can be interpreted as a very weak significance. These significance levels and magnitudes are similar to the Mann-Whitney tests presented earlier. However, recollecting the findings of Plassman et al. (2008), the

limited explanatory power immediately spurs speculation that a stronger price signal would have led to stronger significance, but that is beyond the reach of this analysis. To continue, the F-statistic for column (1) indicates very little joint significance. This is echoed in column (2), where we see that the significance of *group*3 is reduced when left as a single explanatory variable. The alternative formulation in column (3) also shows only limited significance. In column (4) we see a statistically significant effect (t=1.77) from *bothsig*, i.e. of not belonging to the benchmark group, but R^2 does not budge from the 1%-level.

The somewhat vague results warrant a note about the meaning of the t- and F-statistics. As motivated in section 4.1, it is assumed that the null hypothesis for the coefficients on the treatment variables, $\mathrm{H}_0 : \beta = 0$, has a one-sided alternative of the form $\mathrm{H}_\mathrm{a}^1 : \beta > 0$. This is reflected in the reported p-values. An assumption of a two-sided alternative of the form $\mathrm{H}_\mathrm{a}^2 : \beta \neq 0$ would have led to p-values that are twice as large as the ones reported (i.e. less significance). In contrast, the F-statistic for joint significance assumes, by construction, a two-sided alternative, which means that it might underestimate the actual model significance. Nevertheless it is included for completeness. There is also a reason for the drastic increase of the model F-statistic in column (4): with only one explanatory variable the F-statistic is actually the square of the t-statistic, which is illustrated by the fact that the two-sided p-value for $t = 1.77$ is 0.078, which is identical to the model p-value in column (4) (and likewise for the second column).

From one perspective, these results are a little discouraging, since the experiment was conceived in anticipation of a stronger effect from both signals, as well as stronger explanatory power. In addition, the regression results are not particularly helpful in explaining the variation seen in figure 3 above. Therefore it is reasonable to continue the analysis by extending the regressions in table 4 with some additional control variables to see if any other interesting effects can be detected. Let's start by looking at the incentive effect. In table 5, the incentive dummy has been added to the regressions.

Table 5 on the next page corroborates the Mann-Whitney tests of table 3 and contains arguably one of the most interesting findings in the study: as can be seen, the incentive treatment has a strong, statistically significant and positive effect on valuations. Note that the dependent variable is now *ownvaluation2*, as explained in the section on data description. Remember also that the incentive treatment is not independent within sessions, as explained in section 4.3. To compensate for this, the regression in table 5 has been clustered on session date (so t-values are not directly comparable between the two tables, since the degrees of freedom are greatly reduced with clustered standard errors). It is worth to point out, however, that for this particular sample, it turns out that using clustered (robust) standard errors leads to *increased* significance for every variable that is already significant in the unclustered regression. The unclustered regressions still produce highly significant results, but these are not reported since they do not alter the conclusions. This is reassuring in view of a recent paper that cautions against relying too much on robust standard errors when drawing conclusions from data if the clustered and unclustered specifications diverge (King and Roberts, 2012). Before commenting further

Table 5: Adding Incentives
(Clustered Std. Errors)

Regressand: *ownvaluation2*	(1)	(2)	(3)	(4)
group2	6.698			
group3	13.958***	9.333**		
	(2.78)	(1.81)		
group4	7.893***			
	(2.75)			
price			0.407	
story			7.620**	
			(2.44)	
bothsig				9.379**
				(2.61)
incentivized	22.458⋆⋆⋆	22.754⋆⋆⋆	21.951⋆⋆⋆	21.755⋆⋆⋆
	(4.97)	(4.90)	(4.81)	(5.36)
_constant	69.165	73.563	72.296	69.753
R^2	0.066	0.059	0.059	0.061
F-stat	10.528	11.998	13.717	16.623
Prob $> F$	0.001	0.002	0.000	0.000
N	253	253	253	253

$*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$ (One-sided)
$\star$ $p < 0.10$, $\star\star$ $p < 0.05$, $\star\star\star$ $p < 0.01$ (Two-sided)
Selected t-values in parentheses.

on the results, it is timely to point out that table 5 reports both one-sided and two-sided p-values. Since there is no obvious *a priori* assumption about the sign of the incentive effect, it seems reasonable to assume a two-sided alternative to be on the safe side.

If we look at column (1) in table 5 we see that on average, respondents in the benchmark treatment (i.e. $group2 = group3 = group4 = 0$) who were subject to the incentivized treatment increase their valuation of the wine by as much as 32% compared to the control group (= 22.46/69.17), and that the result is significant at the 1%-level (in fact, the p-value is only 0.05%). The overall model significance also increases, and although still relatively modest at around 6%, R^2 is no longer negligible. We also see that the treatment effects become markedly more statistically significant in all columns compared to table 4. In sum, table 5 suggests that there is indeed a signal treatment effect for group 3 and 4 when incentives are added to the equation. The significant coefficient on *group4* should be seen against the lack of significance for *group2*, since this hints towards the possibility of a complementary effect between the monetary and non-monetary signal, even when the monetary signal in itself is not significant (compare with *price* and *story*

in column (3)). Against this setting we will add control variables and check for potential interaction effects between treatments and incentives, but before we proceed, table 5 hides one additional curiosity that deserves mention.

As noted in section 4.3, the modal result of adding incentives in experimental studies is to reduce variation in results. In this data sample the effect is the opposite, as illustrated by table 6. On the one hand, it is necessary to consider that for *ownvalution2*, some observations have been assigned values at the lower bound of the valuation scale in the questionnaire, which obviously reduces variation, but on the other hand, if non-incentivized respondents had also marked their reply on such a scale, presumably that's exactly where they would have put their mark (or, similarly, some incentivized respondents would presumably have indicated a lower value than 40 or 50 if they had been allowed to do so). In any case, the effect is present *even* in the non-constrained data ($incentivized = 0$) of the variable *ownvaluation*. A variance ratio test for *ownvalution2* grouped by *incentivized* gives an $F(41, 210)$-statistic of 0.43, which is significantly different from one at the 0.2%-level. A similar test for *ownvalution* gives an F-statistic of 0.74, which is significantly lower than one (assuming a one-sided alternative) at the 13%-level. Clearly then, in this particular setting, incentives increase variability in outcomes.

Table 6: Variance Comparison for Incentive Effect

	incentivized	Obs.	Mean	Std. Dev.	95% Conf. Int.
ownvaluation	0	42	71.67	32.72	61.47–81.86
ownvaluation	1	211	98.18	37.91	93.03–103.32
*ownvaluation*2	0	42	76.45	24.99	68.66–84.24
*othervaluation*2	1	211	98.18	37.91	93.03–103.32

5.2 Extensions to Baseline Regressions

This section presents selected additional regression tables to complement the main findings in the previous section. To avoid cluttering the exposition too much, some tables have been confined to appendix A.

Adding Control Variables

Table 7 on page 32 adds a substantial set of controls to check the robustness of the results in tables 4 and 5. Note that table 7 does not report model F-statistics. The reason is that there are not enough degrees of freedom after the data is clustered, due to the many variables. However, in the unclustered version of this regression (not reported, but again similar in conclusions to the clustered specification), the overall model significance

is at least 0.01% in column (1)–(4). Note also that R^2 has been replaced with its adjusted counterpart to compensate for potential over-specification, which reduces R^2 by approximately one-half.

The inclusion of the control variables reduces the significance of the non-monetary signal represented by *group*3 and *story*, both statistically and in magnitude, although the t-value of 1.34 for *group3* in column (1) is only 0.4 percentage points away from being statistically significant at the 10%-level. However, the non-monetary signal associated with group 3 fails to pick up any significance when left alone, as in column (2). Likewise, *story* does not show any clear significance ($p = 15\%$) in column (3), which leaves out the interaction effect of the two signals. In contrast, the coefficient on *group*4 remains statistically significant ($p = 3.7\%$) and keeps roughly the same magnitude as in table 5. The incentive effect also remains statistically significant in all specifications, but the magnitude is now only half of that reported in table 5.

Furthermore, some of the controls pick up statistical significance. It is not particularly surprising that self-reported knowledge about wine (*know*), or the amount of money usually spent on wine (*usualprice*) can affect valuations. As mentioned in the beginning of section 5, *usualprice* is a candidate determinant of *ownvaluation*. However, the overall significance (adjusted R^2) with the controls included is still only around 10%, and, as shown by the R^2-value in the extra column (5) of table 7, *usualprice* on its own only explains around 2% of the total variation in *ownvaluation2*, so something else is clearly going on.

It is more challenging to explain why the group of respondents who have a Master's degree (*master*) should stand out as so statistically dominant. Could this be a sign of some hidden interactions? The only obvious candidate for such an effect would be an interaction term between age and education level. But before looking for such an effect, remember that it was hypothesized in section 4.4 that age might have a negative effect on results. The results from table 7 do not provide any overwhelming support for that hypothesis, but at least the sign on *age* is negative, and—to squeeze out significance—if we remove the clustered standard errors and allow for a one-sided alternative of the form $\mathrm{H_a} : \beta < 0$, then the coefficient on age becomes significant at levels around 12%–13%, depending on the specification (not reported in any table). In any case, it seems reasonable to check for interactions between *age* and various variables that are related to education. However, adding such interactions (excluded for brevity) with variables *know*, *highschool*, *bach*, *master* and *phd* fails to alter the significance pattern, so for now we simply have to accept that, *ceteris paribus*, becoming a Master's graduate reduces one's fondness of Georgian-made, skin-macerated white wines. A bad omen for the author.

Table 7: Adding Controls
(Clustered Std. Errors)

Regressand: *ownvaluation2*	(1)	(2)	(3)	(4)	(5)
group2	6.227				
	(1.25)				
group3	5.845	1.466			
	(1.34)	(0.28)			
group4	7.471**				
	(1.98)	(0.26)			
price			3.996		
story			3.535		
bothsig				6.621**	
				(2.45)	
incentivized	13.088⋆⋆	13.295⋆⋆	12.928⋆	13.204⋆⋆	
	(2.30)	(2.26)	(2.27)	(2.23)	
age	-0.278	-0.272	-0.284	-0.274	
income	0.018	0.002	0.021	0.014	
male	6.070	6.229	6.113	6.060	
wineeduc	11.992	11.713	11.668	12.109	
know	-3.691⋆⋆	-3.750⋆⋆	-3.650⋆⋆	-3.719⋆⋆	
highschool	9.010	8.641	8.573	9.187	
bach	0.015	0.409	-0.057	0.098	
master	-14.594⋆⋆	-14.370⋆⋆	-14.780⋆⋆	-14.467⋆⋆	
phd	-3.131	-2.486	-3.268	-2.982	
weekdrink	8.353	8.957	8.283	8.449	
weekbuy	-3.333	-3.477	-3.070	-3.493	
usualprice	0.217⋆⋆	0.223⋆⋆	0.221⋆⋆	0.216⋆⋆	0.184⋆⋆
constant	74.915	78.211	75.700	74.877	76.075
Adj. R^2	0.111	0.112	0.114	0.119	0.021
N	232	232	232	232	237

$*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$ (One-sided)
$\star$ $p < 0.10$, $\star\star$ $p < 0.05$, $\star\star\star$ $p < 0.01$ (Two-sided)
Selected t-values in parentheses.

Treatment Interaction Effects

Let's know turn the attention to possible interactions between the signal treatments and the incentive treatment. The regressions are shown in table 9 on page 46 in the appendix (note that the variable *incentivized* has been abbreviated as '*inc*', not to be confused with

income). While incentives are still highly significant of themselves, statistical significance also appears in the interaction terms between the incentive treatment and both of the signal treatments. The effect is strongest for the non-monetary treatment (captured by $inc*group3$ and $inc*story$). For example, in column (1) of table 9 we see that belonging to treatment group 3 and being incentivized now has a predicted value of $109.65 = 72.00 - 2.00 + 19.07 + 20.58$, but that belonging to group 4 and not being incentivized gives a value of $81.00 = 72.00 + 9.00$ (compare with the Mann-Whitney tests on page 26). Note also that, as in the case without interactions, the significance of the interaction terms decreases if the same regressions are run without clustered standard errors (not reported). In any case, the results support an interpretation of an interaction term between treatments, and solidify the notion that incentives by themselves are significant.

Further Robustness Checks

As discussed in section 4.5 about the possible inflation effects, we can use *adjval* as an alternative to *ownvaluation* as a robustness check. Doing so does not alter the results already presented: all signal treatments and the incentive treatment are significant at the 5%-level, but the details are excluded to save space. It was also noted in the section on a possible inflation effect that we could run regressions on subsamples as an extra check. It turns out that such an analysis throws around the significance between the treatment groups depending on the specification, and this is most likely because there are only 20–30 observations in each treatment group in the restricted samples. The details are not presented.

A further extension to the baseline specifications is to use proxies for *ownvaluation*. Table 10 in appendix A, page 47, explores some options. The first three columns of that table use variables that were included in the survey questionnaire primarily in order to function as such proxies (refer to table 13, summary statistics, for a reminder). Neither signal treatments nor the incentive treatment show any statistical significance at all in those three columns. The fourth column has $d\_usual = ownvaluation - usualprice$ as dependent variable, and here the signaling effects remain statistically significant, indicating that respondents are induced by the signals to value the wine higher than they typically spend on wine. Again, the significance of treatment 4 suggests a complementary effect of the signals. The incentive effect is not significant, however. The lack of significance of the incentive treatment is hardly suprising; indeed, d_usual is a linear function of *usualprice*, which should not be correlated with the quasi-random *incentivized* dummy. On the other hand, the lack of any treatment effect on the non-monetary proxies can be given a theoretical interpretation, since it suggests that it is not primarily the perception of taste or quality that is altered by the treatments, but rather the valuation directly. To put it more succinctly: this result shows that there is no linear relationship between quality, taste and value in the setup of this experiment.

However, an interesting pattern nevertheless occurs if we look at the distributions of the variables for quality and taste. As shown in figure 4 (p. 34), the distribution of reported

taste is quite different from the other three variables, with a higher standard deviation. That the standard deviations are statistically different is also confirmed by a variance ratio test: the difference between *owntaste* and *ownqlty* is significant at the 1.9%-level ($F = 1.32$, with 285 degrees for freedom in both variables). Also, although maybe not obvious from the figure, there is a statistically significant difference in standard deviation between *ownqlty* and *otherqlty* ($p = 0.5\%$). Finally, it can be noted that the mean of *owntaste* is around 6% lower than *ownqlty* ($p = 1\%$). These results might be a little surprising to those who would expect that on average, people in general would not make a difference in their ratings of taste versus quality, or for that matter their own tastes relative to the general public.

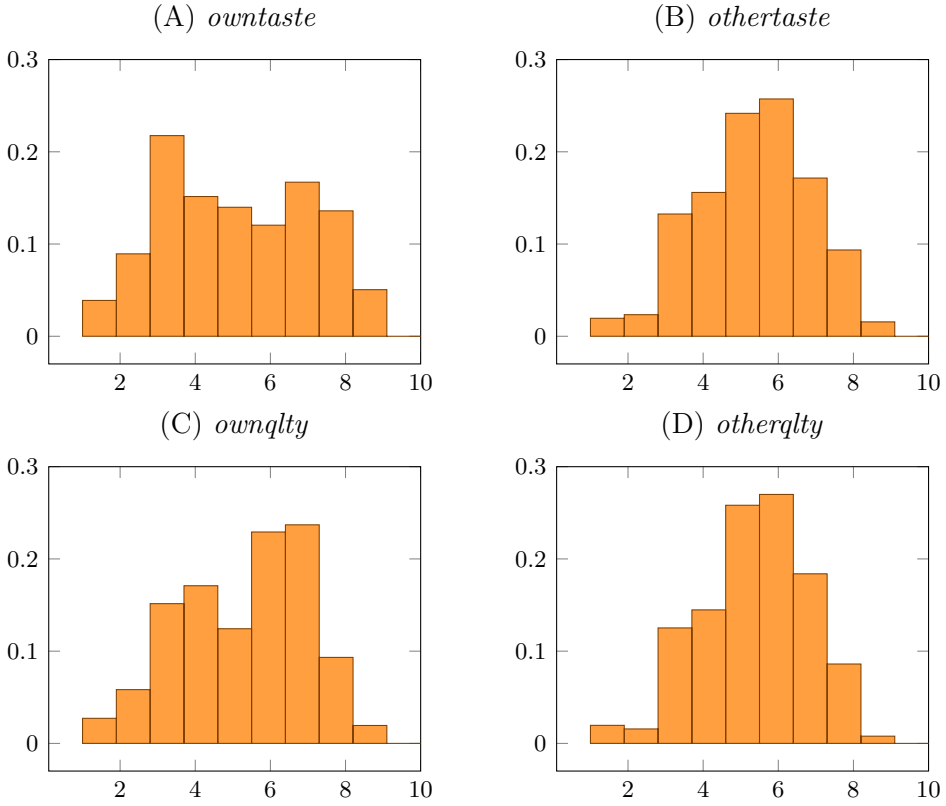

Figure 4: Distributions of Quality and Taste

Table 11 (appendix A, page 48) briefly shows how word count and duration for each respondent depend on a set of controls, as a proxy for how engaged the respondent becomes in the tasting. Apart from the unsurprising effect on word count from being knowledgable about wine, or the effect on duration from being a wine professional, this table also tells us that the incentive effect has a positive effect on both those variables as well, suggesting that more mental resources are devoted to the evaluation process if there is a financial incentive involved. In particular, being incentivized increases the duration by almost 50% (= 2.86/5.98) at a two-sided significance level of 1%. However,

these are necessarily very crude measurements. Duration is essentially calculated from the respondents' self-reported start and end time, and word count can depend on unobserved variability among respondents, such as whether an individual is fond of writing or not. In addition, some of the increased duration might be because incentivized respondents take longer to understand the valuation question, although those instructions were given before the start of the tasting.

To round off the data exhibition, table 12 on page 49 in the appendix shows what happens when respondents' assessment of others' valuation of the wine is used as the dependent variable (*othervaluation*). Now, for the first time thus far, the monetary signal picks up some significance at the 5%-level, as seen in the first and third columns of that table, while the non-monetary signal does not. Note also in column (1) that the interaction term between the two signals (*group4*) remains significant at the 10%-level, indicating that a complementary effect might be at play in this setting as well (compare for example with the effect on *group4* in table 7 above, where *ownvaluation2* is the dependent variable). The statistical significance of *price* on *othervaluation* is interesting, since it indicates a possible connection with social considerations, as hypothesized in section 3.3.

Of course, the standard economic interpretation of this price effect would be to say that it is only an illustration of the market's coordination mechanism. This might indeed be the most plausible explanation, but for two circumstantial empirical facts it might be overly simplistic to neglect the possibility of other mechanisms lurking beneath: firstly, the market coordination argument might work fine here, but yields little insight in the related case of Plassman et al. (2008) where a causal effect of prices on *perception* was established. Secondly, if the coordination mechanism were indeed performing its function diligently, then the mean outcome should not only coordinate closer to the actual market price, but the variance should also decrease. However, this is not the case, as shown in figure 5, which illustrates the sample distributions of *othervaluation* for the whole sample (A) and for the subsample who received the price signal (B). Note that histogram (A) is the same as histogram (C) in figure 3 (page 24). As can be seen, the distributions are very similar, and a variance ratio $F(145, 137)$-test comparing the group with $price = 1$ against the group with $price = 0$ yields an F-statistic of 1.0035, suggesting strongly that the standard deviations are equal in the two groups, contrary to what one might have expected intuitively.

As we saw in figure 3, *othervaluation* has a lower variance than *ownvaluation*, but the variance could still be driven partly by the same effect for the two variables. The likely explanation is of course that respondents taste a wine that is very unfamiliar. Some like it, some do not, and given the result reported here, the personal sensory experience is probably a much stronger determinant for respondents' expectations about other people's perception than the treatment signals, including the effect of having received information about the market price. Maybe a reasonable way to end the results section and bridge to the discussion is then to note that sometimes sensory experience trumps the price as a coordination mechanism, a finding that contrasts with the opposite effect documentet by Veale and Quester (2008) when the wine was from a familiar category.

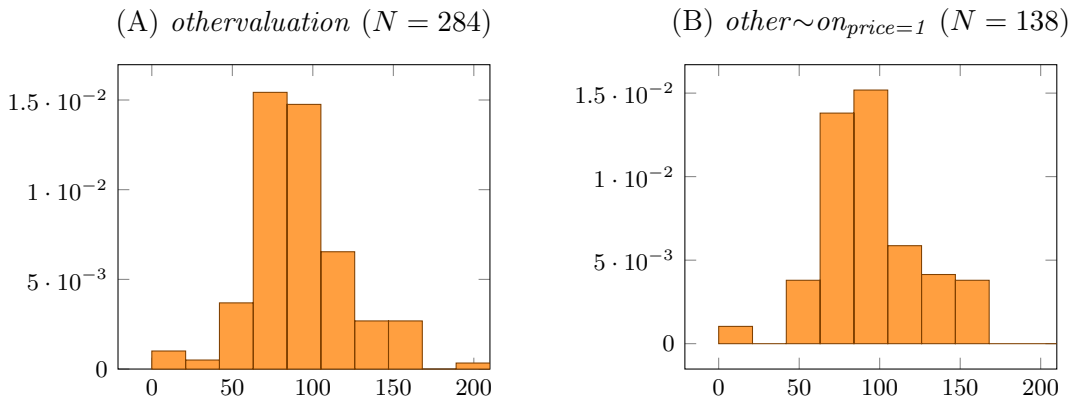

Figure 5: Distribution of Assessment of Others' Valuation

6 Discussion

This section relates the findings from the previous section to the theory and literature presented earlier, and also discusses some aspects of generalizability and statistical inference in view of the empirical results.

6.1 Relating the Results to Theory and Literature

Earlier studies have shown that revealing the price of a wine can affect the perception positively if the price is high. Plassman et al. (2008) showed that a strong price signal (albeit a false one) could alter the perception of a wine to the point that it affected the respondent's brain activity. This result certainly challenges the basic idea of rational choice theory that preference reversals do not occur. Also, it suggests a strong feedback mechanism between price and perception, which in turn suggests feedback loops between prices, preferences and expectations. To account for such effects in formal models would require that the standard downward sloping demand equation, which is linear in price, should be replaced with a set of coupled differential equations, a challenge that would complicate economic analysis tremendously.[7]

In this study, the price communicated to respondents was around 50 % larger than what the average respondent typically pays for a bottle of wine, but it was still within a fairly normal price range (within 1.5 standard deviations from the mean). This contrasts with Plassman et al. (2008), where the price was increased by an order of magnitude. Perhaps the difference in results suggests that a price signal needs to be very strong,

[7]The complications involved have not prevented brave attempts to formulate such models, as shown for example by Englezos and Karatzas (2009). However, although the technical feasibility of such models is interesting in itself, it does not provide much insight into the everyday aspect of decision making which is the focus here.

several standard deviations, in order to trigger any mechanism that would directly affect expectations, at least when the product is unfamiliar. Also, the limited effect of the signal treatments is in line with the findings of Levin and Gaeth (1988) in that framing effects decrease when participants actually taste the product.

The three main empirical results documented by the analysis are: (1) significant variability in subjective valuations and tastes, (2) a limited signal treatment effect, and (3) a significant effect on valuations from incentives, both in terms of increasing valuations, but also by increasing the variance of valuations. In contrast with Plassman et al. (2008) and Almenberg and Dreber (2010) who identify an effect of price on perception, this study finds an effect of incentives and signals on valuation, but no effect directly (or via valuation) on perception, as measured by self-reported taste experience. There seems to be a missing link for a general understanding these complex phenomena. A tentative hypothesis would be to try to fill that gap by considering how social considerations enter into the equation when the consumed product does not match expectations.

The lack of any strong signal treatment effect can be interpreted as a justification for rational choice theory: preferences seem to be robust even to moderate signals designed specifically to influence them. In defense against such a view it could be tempting to simply dismiss the experimental design and argue that the treatment effect is all about the quality of the signal, and that stronger signals would have found a stronger effect. But tempting though it may be, deferring all explanatory power to a context-dependent signal would be to stretch a point, and it is not particularly insightful. As was mentioned in the introduction, the relevance of rational choice depends on the research question. Indeed, based on the regression results from this analysis, we will remain hopelessly ignorant if our ambition is to predict valuation outcomes for a certain individual who tastes our wine, so in that respect rational choice underperforms. On the other hand, the rigidity of preferences to manipulation and the strong incentive effect fit squarely into a rational choice framework. Also, as illustrated by Veale and Quester (2008), conjoint analysis, which puts preference ranking at the center stage in the analysis of wine perception, can be informative in, for example, the strategic considerations of a firm about where to invest its capital.

The framing effect from the treatments, although limited, and the variability in valuations and tastes, are easier to understand from a heuristic perspective. For example, the lack of a cognitive reference frame could help to explain the increased variability. The apparent but weak effect detected here by the non-monetary treatment could be interpreted as a fulfilling an anchoring function in providing a justification to respondents for the strange taste. It would be informative to conduct a similar experiment with a wine that is chosen to *match* expectations rather than upset them to see if the treatment effect becomes stronger. One casual observation in support of this hypothesis is that many longtime consumers of skin-macerated white wines, including this author (thereby mimicking the start of his coffee-consumption career), confess to having had doubts about the taste initially, but that the affection increased with time as it became more familiar. This hypothesis supports the long-standing notion in economics that preferences are guided

by habits, as mentioned in section 2.1. The results presented here, although inconclusive on that note, are certainly consistent with such an interpretation.

The interpretation of the causation of such a habit effect actually depends on whether we adopt the rational or the heuristic paradigm: The rational view would understand the person as arguing along the lines of "I know how much I usually pay for a bottle of wine, and that's how much I'm willing to pay today as well," making the choice conditional on known facts. The heuristic view on the other hand would regard a person as thinking more like "I remember that wines typically cost so much, so the price of this bottle is probably also in the same ballpark." This distinction between the rational and heuristic paradigm is in line with the view of Kahneman (2011), that people can be thought of as having (at least) two fundamentally different thought processes, one fast-thinking, heuristically guided, and one slow-thinking process with room for calculations that require effort. The effect of incentives documented here could potentially be understood as a trigger of certain cognitive processes. Furthermore, a person with interest and knowledge about wine can, through past experience, internalize known facts about prices and wines into the fast-thinking brain track. Such a person would then be able to rely implicitly on rational considerations in the choice process. In contrast, the heuristic approach is illustrated by the anonymous wine survey respondent quoted in the introduction, who is not particularly fond of the taste, but consciously confesses to not having the capacity to make a rational judgement about the wine's quality. This observation is best summarized in figure 4 (p. 34), where we saw a statistical difference between self-reported perception of quality and taste, but also between own taste and the assessment of others' taste. This ambiguity should become particularly relevant in the valuation process when the actual tasting experience does not fit the respondent's reference frame, and it is only natural then that the variability in perception across respondents is quite significant.

6.2 Methodological Limitations

The highly context-focused approach adopted in this study gives particular reason to scrutinize the statistical validity. Although random treatments bring the study closer to the experimental ideal, the setup is not fool-proof. As argued above, while the signal treatments are randomized at the individual level, the incentive treatment only varies across sessions. In fact, all the non-incentivized sessions where performed either before or after the incentivized sessions (although the idea of incorporating incentives was central from the conception). The experience gathered at the first non-incentivized sessions helped inform the design of the incentive-compatible experiment, and, in all honesty, the very last sessions (non-incentivized) were conducted *in the hope* of finding an effect, *after* it became clear that the variation within the incentivized treatment was not as strong as expected. It seems inevitable then, that certain subtleties in how the experiment is framed to participants, evoked by unconscious researcher bias, might drive results in a non-trivial way. This is a general problem in the social sciences, and the effect might be accentuated by a publication bias, i.e. the tendency that journals only publish positive

results and reject insignificant results though insignificance itself might be important for the general theory. Brian Nosek has offered a compelling argument for why such considerations deserve to be taken seriously in all social sciences, and presents some ideas and practices for how to try to avoid such systematic biases (Nosek, 2012), for example by enabling researchers to make hypotheses public before the data gathering is initiated, thereby limiting 'researcher degrees of freedom.'

For this particular study, even if one would assume that the pitfalls outlined in the previous paragraph were avoided, the possibility remains that a few sessions are driving results. In defense against that possibility, it is worth to point out, however, that there is no obvious difference between the types of participants in the various sessions. For instance, an incentivized session was conducted with cadets at a military college, and a non-incentivized session among staff and students at the National Defense Academy. Likewise, both non-incentivized and incentivized sessions were conducted with wine clubs and sommelier students, and also in office environments, where selection bias might otherwise have caused concern. One possible culprit, however, could be the number of participants in a single session. The lottery aspect of the incentive treatment necessarily required that several people participate in one session (typically 15–30 participants), whereas the non-incentivized sessions could easily be conducted with only a handful of people. In practice, when the word 'session' has been used throughout the text, it actually means "session or set of mini sessions at one location on a particular date," but the data gathering process offers no possibility of grouping participants in smaller clusters than session date (partly due to concerns about anonymity for respondents). On that note, it is time to summarize and conclude this study.

7 Conclusion

The assumption of a rigid preference structure which is central to rational choice theory stands up pretty well to the test conducted here. In this particular case, the considerable variability shows that personal taste clearly trumps signals designed to lure the respondent to change his or her perception in a favorable direction, even in the presence of financial incentives. However, the results are somewhat ambiguous since a limited treatment effect is nevertheless identified in the data analysis, suggesting that the experiment does indeed approach the boundaries of the domain of rational choice.

One hypothesis spurred out of previous research was that there could be an interaction effect between the monetary and non-monetary signal, which inspired the design of the experiment as well as the title of the thesis. The inclusion of treatment group 4 was intended to capture such an interaction term, and, as has been shown, the effect indeed appears to exist in the data. The empirical results are perhaps not overwhelming evidence to make any definitive conclusions, but the effect nevertheless appears to be statistically significant even after robustness checks. Thus, the price signal and the conceptual signal appear to be complements rather than substitutes in this context.

On the other hand, the results indicate an unambiguous, positive effect of financial incentives on valuations (and on cognitive measures like word count and duration), a result which is neither intuitive nor expected given the standard view in the literature. As noted in the introduction, any effect of incentives on preferences formation is a motivation for economists to incorporate such considerations into the analysis, and not only leave the work to psychologists.

The choice of method and experiment design was in part motivated by an attempt to avoid 'contaminating' respondents by an overly enthusiastic experimenter with a personal stake in the financial success of the wine used in the tastings. With hindsight, maybe this caution was emphasized to the point of blurring potentially interesting results that could have emerged if the information was communicated to participants more like a real sales pitch, an idea with some casual support from a recent review of 'orange wines,' as they are sometimes called, including the one used in this experiment, in *The Wine Enthusiast Magazine*: "To the uninitiated, their initial charm was their challenging nature ... You couldn't sell them the way you do normal wines. If you put a big skull and dagger on them, or warning signs telling people they can't handle them, they sell out the roof" (Iijima, 2013). Nevertheless, the reader hopefully agrees that the weak effect of the various treatments confirm that the method was successful in mitigating the obvious risk of researcher bias.

It is also hoped that the reader finds that this thesis has managed to argue credibly that prices deserve to be regarded as more than just a market-clearing mechanism, even in formal economic modeling. The notion that prices convey stories should inspire the researcher to sometimes take a step back and consider our understanding of the wider context, to which economic models ultimately aspire to make a relevant contribution. On the theoretical front, and also in relation to how economics is taught, any effort to clearly articulate the boundaries of rational choice theory should be welcomed and help to enrich our view of human behavior within an economic framework.

To sum up, the results illustrate two points that can be related to the theories presented in section 2. Firstly, the strong impact on valuations of incentivizing respondents underscores the pivotal insight of economic theory in general, and rational choice theory in particular, that incentives matter for economic outcomes. Secondly, the limited but nonetheless detected effect from the signal treatments and the great variability in valuations can be understood as an illustration of the cognitive framing mechanisms that are central to the heuristic approach emphasized by behavioral economics and psychology. As for the theoretical contribution from neuroscience, it remains for future research to show whether perception and valuation in wine tasting can actually be predicted at the moment of sensory experience, but recent advances, as illustrated by Salimpoor et al. (2013), certainly render such a hypothesis plausible. The overlap between sensory experience and social interactions seems to be a particularly exciting field of research, with some low-hanging grapes still available to be fermented.

References

Allison, R. I. and K. P. Uhl, "Influence of Beer Brand Identification on Taste Perception," *Journal of Marketing Research* 1 (1964), 36–39.

Almenberg, J. and A. Dreber, "When Does the Price Affect the Taste? Results from a Wine Experiment," SSE/EFI Working Paper Series in Economics and Finance 717, Stockholm School of Economics / The Economic Research Institute, March 2010.

Angrist, J. D. and J.-S. Pischke, *Mostly Harmless Econometrics. An Empiricist's Companion* (Princeton University Press, 2009).

Beckert, J., "Where Do Prices Come From? Sociological Approaches to Price Formation," *Socio-Economic Review* 9 (2011), 757–786.

Bentham, J., *An Introduction to the Principle of Morals and Legislations* (Reprint, Oxford, UK: Blackwell [1948], 1789).

Bowles, S. and S. Polanía-Reyes, "Economic Incentives and Social Preferences: Substitutes or Complements?," *Journal of Economic Literature* 50 (2012), 368–425.

Camerer, C. F. and R. M. Hogarth, "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Producion Framework," *Journal of Risk and Uncertainty* 19 (1999), 7–42.

Chow, G., "Rational Versus Adaptive Expectations in Present Value Models," *The Review of Economics and Statistics* 71 (August 1989), 376–384.

Darlow, A. L. and S. A. Sloman, "Two Systems of Reasoning: Architecture and Relation to Emotion," *Wiley Interdisciplinary Reviews: Cognitive Science* 1 (May/June 2010), 382–392.

Deaton, A., "Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development," The Keynes Lecture, British Academy, 9 October, 2008, Research Program in Development Studies, Center for Health and Wellbeing, Princeton University, January 2009.

Debreu, G., "Representation of a Preference Ordering by a Numerical Function," in R. M. Thrall, R. M. Coombs and C. H. Davis, eds., *Decision Processes* (New York: John Wiley, 1954), 159–165.

Deliza, R. and H. J. H. McFie, "The Generation of Sensory Expectation by External Cues and its Effect on Sensory Perception and Hedonic Ratings: A Review," *Journal of Sensory Studies* 11 (1996), 103–128.

Eaton, C. B. and M. Eswaran, "Well-being and Affluence in the Presence of a Veblen Good," *The Economic Journal* 119 (2009), 1088–1104.

Ellingsen, T., B. Herrman, M. A. Nowak, D. G. Rand and C. E. Tarnita, "Civic Capital in Two Cultures: The Nature of Cooperation in Romania and USA," Working paper, 1 November 2012, available at SSRN: http://ssrn.com/abstract=2179575.

Englezos, N. and I. Karatzas, "Utility Maximization with Habit Formation: Dynamic Programming and Stochastic PDEs," *SIAM Journal on Control and Optimization* 48 (2009), 481–520.

Gilovich, T., D. Griffin and D. Kahneman, eds., *Heuristics and Biases: The Psychology of Intuitive Judgement* (Cambridge University Press, 2002).

Gorman, W. M. T., "Tastes, Habits and Choices," *International Economic Review* 8 (June 1967), 218–222.

Green, P. E. and V. Srinivasan, "Conjoint Analysis in Consumer Research," *Journal of Consumer Research* 5 (September 1978), 103–23.

———, "Conjoint Analysis in Marketing: New Developments With Implication for Research and Practice," *Journal of Marketing* 54 (October 1990), 3–19.

Harrison, G. W. and J. A. List, "Field Experiments," *Journal of Economic Literature* XLII (December 2004), 1009–1055.

Hirsch, F., *The Social Limits to Growth* (London: Routledge & Kegan Paul, 1977).

Hoch, S. J. and Y.-W. Ha, "Consumer Learning: Advertising and the Ambiguity of Product Experience," *Journal of Consumer Research* 13 (September 1986), 221–233.

Iijima, A. L. C., "Orange Wines Will Never Be Mainstream. But A Case for Why They're More Than a Dying Trend," *The Wine Enthusiast Magazine* (5 March 2013), online publication: http://www.winemag.com/.

Jehle, G. A. and P. J. Reny, *Advanced Microeconomic Theory*, third edition (Essex: Prentice Hall, 2011).

Johnson, H., *The Story of Wine — New Illustrated Edition* (London: Mitchell Beazley, 2004), first published in 1989 as *The Story of Wine*.

Kahneman, D., "Preface," in Kahneman and Tversky (2000), ix–xvii.

———, *Thinking, Fast and Slow* (New York: Farrar Straus Giroux, 2011).

Kahneman, D. and A. Tversky, eds., *Choices, Values and Frames* (Cambridge University Press, 2000).

Kahneman, D., A. Tversky and P. Slovic, eds., *Judgement under Uncertainty: Heuristics and Biases* (Cambridge University Press, 1982).

Kahneman, D., P. P. Wakker and R. Sarin, "Back to Bentham? Explorations of Experienced Utility," *Quarterly Journal of Economics* 112 (May 1997), 375–405.

King, G. and M. Roberts, "How Robust Standard Errors Expose Methodological Problems They Do Not Fix," mimeo, available at http://gking.harvard.edu/files/robust.pdf, September 2012.

Klaaren, K. J., S. D. Hodges and T. D. Wilson, "The Role of Affective Expectations in Subjective Experience and Decision-Making," *Social Cognition* 12 (1994), 77–101.

Lee, L., S. Frederick and D. Ariely, "Try It, You'll Like It. The Influence of Expectation, Consumption, and Revelation of Preferences for Beer," *Psychological Science* 17 (2006), 1054–1058.

Levin, I. P. and G. J. Gaeth, "How Consumers Are Affected by the Framing of Attribute Information Before and After Consuming the Product," *Journal of Consumer Research* 15 (December 1988), 374–378.

McClure, S., J. Li, D. Tomlin, K. S. Cypert, L. M. Montague and P. R. Montague, "Neural correlates of Behavioral Preference for Culturally Familiar Drinks," *Neuron* 44 (14 October 2004), 379–387.

McGovern, P. E., *Ancient Wine. The Search for the Origins of Viniculture* (Princeton University Press, 2003).

Migone, A. and M. Howlett, "Comparative Networks and Clusters in the Wine Industry," AAWE Working Paper 62, American Association of Wine Economists, July 2010.

Mitchell, H., "Knowledge Sharing—The Value of Storytelling," *International Journal of Organizational Behaviour* 9 (2005), 632–641.

Nosek, B., "Nosek on Truth, Science and Academic Incentives," Econtalk podcast, Library of Economics and Liberty, Liberty Fund, Inc., Indianapolis, 10 September 2012.

Osborne, M. J., *An Introduction to Game Theory*, international edition (New York: Oxford University Press, 2009).

Ostrom, E., "Collective Action and the Evolution of Social Norms," *The Journal of Economic Perspectives* 14 (Summer 2000), 137–158.

Plassman, H., J. O'Doherty, B. Shiv and A. Rangel, "Marketing Actions Can Modulate Neural Representations of Experienced Pleasantness," *Proceedings of the National Academy of Sciences of the United Statets of America* 105 (22 January 2008), 1050–1054.

Pollak, R. A., "Habit Formation and Dynamic Demand Functions," *Journal of Political Economy* 78 (1970), 745–763.

Rao, A. R. and K. B. Monroe, "The Effect of Price, Brand Name, and Store Name on Buyers' Perception of Product Quality: An Integrative View," *Journal of Marketing Research* XXVI (August 1989), 351–357.

Robinson, J., "10 Grapes You Must Try: Britain's Veteran Wine Critic and Co-author of a New Book on the World's 1,368 Grapes, Picks the 10 that Most Amazed Her," *Decanter* (7 November 2012).

RVF, 2011, "Dégustation: Découvrez le Goût du Vin en Amphores. 50 Cuvées pour Illustrer une Résurgence," *La Revue du Vin de France* (Novembre 2011), 78–83.

Salimpoor, V. N., I. van den Bosch, A. Randal McIntosh, A. Dagher and R. J. Zatorre, "Interactions Between the Nucleus Accumbens and Auditory Cortices Predict Music Reward Value," *Science* 340 (12 April 2013), 216–219.

Schelling, T. C., "Self-Command in Practice, in Policy and in a Theory of Rational Choice," *American Economic Review* 74 (May 1984), 1–11.

Sloman, S. A. and L. J. Rips, "Similarity as an Explanatory Construct," *Cognition* 65 (January 1998), 87–101.

Thaler, R., "Mental Accounting and Consumer Choice," *Marketing Science* 4 (Summer 1985), 199–214.

Tusche, A., S. Bode and J.-D. Haynes, "Neural Responses to Unattended Products Predict Later Consumer Choices," *The Journal of Neuroscience* 30 (June 2010), 8024–8031.

Veale, R. and P. Quester, "Consumer Sensory Evaluations of Wine Quality: The Respective Influence of Price and Country of Origin," *Journal of Wine Economics* 3 (2008), 10–29.

Veblen, T. B., *The Theory of the Leisure Class. An Economic Study of Institutions* (London: Macmillan Publishers, 1899).

von Mises, L., "Economic Calculation in the Socialist Commonwealth," in F. A. Hayek, ed., *Collectivist Economic Planning* (London: Routledge & Sonds, 1935), 87–130, originally published as "Die Wirtschaftsrechnung im sozialistischen Gemeinwesen" in the *Archiv für Sozialwissenschaften*, vol. 47 (1920).

Appendix

A Tables of Additional Regression Results

Table 8: Baseline Regressions, A Variation

Regressand: *ownvaluation*	(1)	(2)	(3)	(4)
group2	7.927	7.927		
group3	12.072*	12.072*		
group4	8.376			
	(1.31)			
price $*$ *story*		8.376	-11.623	
		(1.31)	(-1.20)	
price			7.927	
			(1.16)	
story			12.072*	
			(1.77)	
bothsig				9.368*
_constant	87.110	87.110	87.110	87.110
R^2	0.014	0.014	0.014	0.012
F-stat	1.169	1.169	1.169	3.128
Prob $> F$	0.322	0.322	0.322	0.078
N	253	253	253	253

$*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$
t-values in parentheses.
See page 26 in section 5.1 for comments.

Table 9: Treatment Interactions
(Clustered Std. Errors)

Regressand: *ownvaluation2*	(1)	(2)	(3)	(4)
group2	16.500			
	(1.55)			
group3	-2.000	-9.345*		
	(-0.47)	(-2.06)		
group4	9.000			
	(1.22)			
incentivized	19.066⋆⋆⋆	15.981⋆⋆⋆	21.238⋆⋆⋆	19.066⋆⋆⋆
	(3.55)	(3.18)	(4.37)	(3.58)
inc ∗ *group2*	-11.395			
	(-0.86)			
inc ∗ *group3*	20.577⋆⋆	23.662⋆⋆⋆		
	(2.96)	(3.34)		
inc ∗ *group4*	-1.131			
	(-0.14)			
price			13.609⋆	
			(1.91)	
story			-4.224	
			(-1.08)	
inc ∗ *price*			-16.154⋆	
			(-1.89)	
inc ∗ *story*			14.637⋆⋆	
			(2.87)	
bothsig				6.233
				(0.98)
inc ∗ *bothsig*				3.768
				(0.50)
constant	72.000	79.345	73.156	72.000
Adj. R^2	0.08	0.07	0.07	0.06
N	253	253	253	253

∗ $p < 0.10$, ∗∗ $p < 0.05$, ∗∗∗ $p < 0.01$ (One-sided)
⋆ $p < 0.10$, ⋆⋆ $p < 0.05$, ⋆⋆⋆ $p < 0.01$ (Two-sided)
incentivized abbreviated as *inc*.
See page 32 in section 5.2 for comments.

Table 10: Proxies as Dependent Variable

Regressand:	*inclbuy*	*ownqlty*	*owntaste*	*d_usual*
*group*2	-0.143	0.036	0.016	7.651*
	(-0.31)	(0.18)	(0.06)	(1.52)
*group*3	0.392	0.194	0.186	7.849*
	(1.07)	(0.69)	(0.49)	(1.57)
*group*4	-0.004	0.274	-0.101	8.443**
	(-0.01)	(1.16)	(-0.36)	(1.87)
incentivized	-0.521	-0.319	-0.349	12.452
	(-0.87)	(-1.01)	(-0.90)	(1.16)
age	-0.049⋆⋆	-0.040⋆⋆⋆	-0.036⋆⋆	-0.275
male	0.842⋆⋆	0.195	0.540⋆	7.108
know	-0.223	-0.116	-0.114	-5.175⋆⋆
intr	0.347⋆⋆	0.028	-0.004	1.784
master	-0.599	-0.677⋆	-1.153⋆⋆⋆	-14.672⋆⋆
weekbuy	-0.252	-0.622⋆	-0.288	-3.024
usualprice	0.007	0.001	0.003	-0.804⋆⋆⋆
		⋮		
Insignificant controls excluded for brevity.				
		⋮		
_constant	4.344	6.943	6.599	70.655
Adj. R^2	0.077	0.065	0.085	0.405
N	232	232	232	214

$*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$ (One-sided)
$\star$ $p < 0.10$, $\star\star$ $p < 0.05$, $\star\star\star$ $p < 0.01$ (Two-sided)
See page 33 in section 5.2 for comments.

Table 11: Word Count and Duration

Regressand:	*wc*	*duration*
	⋮	
	Insignificant treatments excluded for brevity.	
	⋮	
incentivized	3.144**	2.859***
age	-0.205**	-0.071***
winepro	2.478	1.402*
know	0.908**	0.186
highschool	-4.867**	-0.781
phd	-3.621*	-0.132
usualprice	0.076**	0.002
	⋮	
	Controls excluded for brevity.	
	⋮	
_constant	4.865	5.980
Adj. R^2	0.26	0.21
N	232	212

$*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$
(Two-sided)
See page 34 in section 5.2 for comments.

Table 12: Others' Valuation

Regressand: *othervaluation*	(1)	(2)	(3)	(4)
group2	7.470**			
	(1.90)			
group3	3.693	-1.541		
	(1.00)	(-0.48)		
group4	8.994*			
	(1.66)			
price			6.434**	
			(1.76)	
story			2.602	
			(0.74)	
bothsig				6.992**
				(1.85)
incentivized	2.164	2.324	2.112	2.557
	(0.81)	(0.84)	(0.78)	(1.07)
age	-0.548⋆⋆	-0.526⋆⋆	-0.550⋆⋆	-0.535⋆⋆
know	-4.328⋆⋆⋆	-4.644⋆⋆⋆	-4.268⋆⋆⋆	-4.596⋆⋆⋆
master	-12.889⋆	-12.461⋆	-12.993⋆	-12.403⋆
		⋮		
Insignificant controls excluded for brevity.				
		⋮		
_constant	104.221	107.507	104.672	103.793
Adj. R^2	0.173	0.162	0.176	0.176
N	231	231	231	231

$*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$ (One-sided)
$\star$ $p < 0.10$, $\star\star$ $p < 0.05$, $\star\star\star$ $p < 0.01$ (Two-sided)
See page 35 in section 5.2 for comments.

Table 13: Summary Statistics

An asterisk (∗) denotes categorical variables, compare w. questionnaire in Appendix B.
Indented variables are derived from corresponding categorical variable.

Variable	Mean	Std. Dev.	Min.	Max.	N	Description
id∗	—	—	1	289	289	Unique identifier
Key variables:						
incentivized	0.84	0.37	0	1	255	= 1 for incentivized respondents
treatment∗	—	—	1	4	289	
benchmark	0.29	0.45	0	1	289	= 1 if treatment = 1
group2	0.23	0.42	0	1	289	= 1 if treatment = 2
group3	0.23	0.42	0	1	289	= 1 if treatment = 3
group4	0.25	0.43	0	1	289	= 1 if treatment = 4. NB: *group4* = *price* ∗ *story*
price	0.48	0.50	0	1	289	= max(*group2*,*group4*)
story	0.48	0.50	0	1	289	= max(*group3*,*group4*)
bothsig	0.71	0.45	0	1	289	= max(*group2*,*group3*,*group4*)
ownvaluation	93.77	38.33	0	210	253	Respondent's valuation of wine
ownvaluation2	94.57	36.94	40	210	253	= *ownval.*, but → 50 if < 50, or → 40 if < 40 and *actualpr* = 139, see section 4.4, page 19.
othervaluation	91.46	30.33	0	200	284	Assessment of other's valuation
Controls (alphabetical order):						
aboveother	0.51	0.50	0	1	252	= 1 if *ownval.* − *otherval.* > 0
aboveusual	0.52	0.50	0	1	219	= 1 if *ownval.* − *usualprice* > 0
actualpr∗	—	—	139	150	289	See sec. 4.5, page 22
adjval	50.1	38.7	-50	160	253	See sec. 4.5, page 22
age	35.39	16.23	18	84	283	Respondent age in years
d_usual	-0.47	45.47	-230	120	219	= *ownval.* − *usualpr.*
dp_other	1.53	32.7	-100	150	252	= *ownval.* − *otherval.*
duration	7.25	2.68	1	15	261	Duration of tasting, min., self-reported
educ∗	—	—	1	10	281	(Some subcategories exlc. for brevity)
highschool	0.09	0.29	0	1	289	= 1 if *educ* = 4
someuniv	0.30	0.46	0	1	289	= 1 if *educ* = 6
bach	0.20	0.40	0	1	289	= 1 if *educ* = 7
master	0.25	0.43	0	1	289	= 1 if *educ* = 8
phd	0.07	0.25	0	1	289	= 1 if *educ* = 10
freqdrink∗	—	—	1	5	285	
drinksev	0.21	0.41	0	1	289	= 1 if *drinkfreq* = 1 (sev. times/week)
drinkweekly	0.39	0.49	0	1	289	= 1 if *drinkfreq* = 2
drinkmonthly	0.34	0.47	0	1	289	= 1 if *drinkfreq* = 3
drinkyearly	0.04	0.20	0	1	289	= 1 if *drinkfreq* = 4
weekdrink	0.61	0.49	0	1	289	= *drinksev* + *drinkweekly*
freqbuy∗	—	—	1	5	285	
buysev	0.01	0.10	0	1	289	= 1 if *buyfreq* = 1 (sev. times/week)
buyweekly	0.20	0.40	0	1	289	= 1 if *buyfreq* = 2
buymonthly	0.58	0.49	0	1	289	= 1 if *buyfreq* = 3

Continued on next page...

... table 13 continued

Variable	Mean	Std. Dev.	Min.	Max.	N	Description
buyyearly	0.15	0.36	0	1	289	= 1 if $buyfreq = 4$
weekbuy	0.21	0.41	0	1	289	= $buysev + buyweekly$
incomegrp∗	—	—	1	12	279	Month. after-tax income, 12 brackets
inc1	14,713	12,726	0	55,000	279	= min. val. of income bracket
income	14.71	12.73	0	55	279	=*inc1*/1,000
intr	6.01	2.22	1	10	285	Self-reported interest in wine
know	4.06	2.04	0	10	285	Self-reported knowledge about wine
male	0.62	0.49	0	1	283	= 1 if respondent is male
otherqlty	5.37	1.61	1	9	284	Assessm't of others' assessm. about qlty.
othertaste	5.34	1.68	1	9	285	Assessm't of others' assessm. about t.
ownqlty	5.27	1.90	1	9	286	Self-reported assm. of wine qlty.
owntaste	5.05	2.18	1	9	286	Self-reported tasting experience
usualprice	95.77	30.52	0	300	267	Amount typ. spent on a btl. of wine
wc	12.34	10.76	0	66	286	# written words in descript. of wine
wineeduc	0.12	0.33	0	1	285	= 1 if resp. has formal wine educ.
winepro	0.10	0.30	0	1	285	= 1 if resp. has wine/food proffession

B Wine Tasting Questionnaire

This appendix complements the information in section 4.1 on page 16. Pages 53–56 illustrate the design of the questionnaire used in the wine tastings, translated from Swedish. The information section carried different pieces of information in the different treatments, as indicated. The design of the questionnaire is therefore of interest for the method employed. Questionnaires were numbered by hand in ascending order at the time of printing. For each printed questionnaire one of the four treatments was randomly assigned. As a result, all four treatments were randomly distributed among participants at each tasting session. The idea was that participants should not realize that they were given different information compared to others. The rationale for this setup was basically experimental logistics, as this turned out to be an efficient way to administer the treatment randomization. Since several of the sessions took place in a social setting, there was an obvious risk of information 'contamination' between respondents. However, some care was taken to avoid this, in particular by instructing participants ahead of the tasting that it was an individual exercise and that they were not expected to communicate. The experimenter also tried to enforce this rule during the actual tasting. Indeed, some participants failed to observe this guideline and spoke about the questionnaire in such a way that others could hear them. However, such behavior was exceptional and not systematic. Furthermore, a significant portion of the sessions took place in a classroom environment where it was easy to ensure that instructions were followed. After each session, the experimenter would typically explain to the participants that they had not received identical information. Reassuringly, the vast majority of participants confessed that they had not realized this during the actual tasting.

The key for eliciting incentive compatible results was the design of question 4 on page 54. For a discussion, see section 4.3 on page 18.

1 INFORMATION

All Treatments

This wine tasting is part of a study for a thesis at university level. All respondents must be at least 18 years old to participate since it involves alcoholic beverages. All collected data is treated anonymously. You will taste a wine and will then be asked to answer some questions. You can continue to taste the wine while you are answering the questions. It is important that you do not communicate with anybody before you have completed the questionnaire and handed it to the tasting administrator. Take the time you need, don't hurry even if other participants should complete the questionnarie before you.

Non-monetary signal: treatments 2 and 4

The wine is called Pheasant's Tears (Swedish: "Fasanens Tårar") and comes from Georgia, a republic located between Russia and Turkey on the eastern shores of the Black Sea. The country is home to the world's oldest finds of winemaking; archeologists have been able to determine that wine was being made in today's Georgia as early as 8.000 years ago, i.e. approximately 3.500 years before the first pyramids were built in Egypt. Both historically and in modern times wine plays a central role in Georgian culture and society. For example, the name Pheasant's Tears comes from a local saying that when a wine has reached perfection, the pheasant cries of joy. Traditional Georgian winemaking only employs organic principles and strives to minimize the human intervention in the winemaking process. Immediately after harvest, the grapes are carefully pressed and the juice is transferred together with the grape skins into a large clay vessel dug into the ground, a so called *qvevri*. A qvevri can contain several tons of wine. These clay vessels are used for several reasons, among other things the method contributes to keeping the temperature stable during fermentation and storage. The wine can stay insulated inside the qvevri for several months before it is time for bottling. A unique trait of the Georgian method is that the grape skins are allowed to remain in the grape juice even in the production of white wines (which is standard for red wines, also internationally). The method leads to darker wines than a typical white wine. During the 20th C. Georgia was part of the Soviet Union and was home to a large and flourishing wine industry, which nonetheless meant that the traditional winemaking was neglected for the benefit of wine production with modern technologies. It is only after economic and political reforms in recent years that several Georgian winemakers have chosen to return instead to traditional methods in the hope of creating an interest for Georgian wines in the rest of the world. The wine that you are about to taste is made according to traditional principles

Monetary signal: treatments 3 and 4

Retail price at Systembolaget [Swedish alcohol retail monopoly]: 139 kronor (standard bottle, 750 ml)

2 TASTE THE WINE

You can continue to taste the wine while you are answering the questions below. Indicate the time when you start tasting (HH:MM): _____ : _____

3 QUESTIONNAIRE

1. How would you describe the wine (Associate freely)

_____
_____
_____
_____
_____
_____

2. How do you rate the quality of the wine on the following scale?

Worst grade ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Best grade

3. How do you rate the tasting experience of the wine?

Worst grade ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Best grade

4. How much would you be willing to pay for a bottle of this wine? – Read the instruction and mark your answer *in krona* on the scale below!

INSTRUCTION: In order to give you an incentive to consider your answer carefully, as a participant you have a chance to win either a bottle of the wine you have tasted or a sum of money. In that case, your answer on this question determines whether you win the money or the wine (see next paragraph). The winner will be drawn randomly after the tasting is finished.

NOTE: If you win, the prize will be determined according to the following procedure: The experimenter will randomly draw one of the numbers between 50 and 210 which have been marked on the scale below. The drawn number is then compared with your mark on the scale. If the drawn number is to the left of your mark, you win a bottle of wine. If the drawn number is to the right of your mark, you win the corresponding amount of money. (If the number coincides exactly with your mark, the prize is also a bottle of wine.) The purpose is to give participants an incentive to answer the question honestly, and thereby simulate a real purchase decision.

Mark your answer *in Swedish krona* on the following scale. The scale has been limited for practical reasons. If your answer does not fit on the scale, pick the closest alternative (50 or 210 kr).

50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210
☐ ☐

(ALTERNATIVE QUESTION 4, only in the non-incentivized version of the experiment: What would you be prepared to pay for a bottle (750 ml) of this wine? _____ kr)

5. How often do you drink wine?

- ☐ Several times per week
- ☐ Once or a few times per week
- ☐ Once or a few times per month
- ☐ Once or a few times per year
- ☐ Never

6. How often do you buy wine?

- ☐ Several times per week
- ☐ Once or a few times per week
- ☐ Once or a few times per month
- ☐ Once or a few times per year
- ☐ Never

7. How much do you usually pay for a bottle (750 ml) of white wine? _____ kr. (Skip this question if you replied "Never" in question 6.

8. How do you think other persons rate the quality of the wine on average?

Worst grade ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Best grade

9. How do you think other persons rate the tasting experience on average?

Worst grade ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Best grade

10. How much do you think other persons are willing to pay for a bottle of this wine on average? _____ kr

11. Do you have any formal education related to wine? Yes ☐ No ☐

12. Do you have a profession that is related to food and wine? Yes ☐ No ☐

13. How do you estimate your knowledge about wine?

Limited knowl. ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Extensive knowl.

14. How would you describe your interest for wine?

Limited interest ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Extensive int.

15. How likely is it that you would buy a bottle of this wine to drink again?

Not likely ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Likely

16. How likely is it that you would tell about this wine to others?

Not likely ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Likely

17. How likely is it that you would buy this wine to serve others?

Not likely ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Likely

18. How big is your monthly income <u>after</u> tax?

- ☐ 0 – 4 999 kr
- ☐ 5 000 – 9 999 kr
- ☐ 10 000 – 14 999 kr
- ☐ 15 000 – 19 999 kr
- ☐ 20 000 – 24 999 kr
- ☐ 25 000 – 29 999 kr
- ☐ 30 000 – 34 999 kr
- ☐ 35 000 – 39 999 kr
- ☐ 40 000 – 44 999 kr
- ☐ 45 000 – 49 999 kr
- ☐ 50 000 – 54 999 kr
- ☐ 55 000 kr or more

19. What education do you have? (Indicate the most accurate alternative)

- ☐ Some elementary school
- ☐ 9 years elementary school
- ☐ Some high school
- ☐ High school degree
- ☐ Vocational degree
- ☐ Some university
- ☐ 3-year university degree
- ☐ 4- or 5-year university degree
- ☐ Some postgraduate education
- ☐ Postgraduate academic degree (research degree)

20. How old are you? _____ years. **21. Sex** Male ☐ Female ☐

Indicate the time when you are finished (hours:minutes): _____ : _____

Thanks for your participation!