

JUDGING THE BOTTLE BY ITS COVER

**WINE LABEL STYLE AND PRICE CUES IN REGULATED
MONOPOLY RETAIL SETTINGS**

EMANUELE DE SANTIS

ELIAS OLIVIER

Bachelor Thesis in Marketing
Stockholm School of Economics
2026



Judging the Bottle by its Cover: Wine Label Style and Price Cues in a Regulated Monopoly Retail Settings

Abstract:

We examine how modern versus traditional wine label design and displayed price shape perceived quality and purchase intention in a Systembolaget-like shelf context. Drawing on visual semiotics, product form theory, and price cue literature, we ask whether a modern label generates higher perceived quality and purchase intention than a traditional one, and whether displayed price moderates the label style effect on purchase intention. We ran a 2x2 between-subjects online experiment with 125 respondents after exclusions. Each participant was randomly assigned to evaluate one fictional wine bottle shown in a Systembolaget shelf scene, varying by label style (modern vs. traditional) and displayed price (99 SEK vs 199 SEK). The manipulations worked as intended, but none of the three hypotheses reached statistical significance. Label style did not significantly affect perceived quality or purchase intention, and the label style x price interaction on purchase intention was not significant either. The clearest finding was that displayed price significantly increased perceived quality, with respondents rating the wine as higher quality at 199 SEK than at 99 SEK. In this regulated retail context, displayed price worked more clearly as a quality cue than label style did.

Keywords:

Wine label design, Perceived quality, Purchase intention, Price cues, Systembolaget, Visual semiotics, Consumer evaluation

Authors:

Elias Olivier (25982)
Emanuele De Santis (25950)

Tutors:

Maja Fors, Affiliated Researcher, Department of Marketing and Strategy

Examiner:

Magnus Söderlund, Professor, Department of Marketing and Strategy

Bachelor Thesis in Marketing
Business and Economics Program,
Stockholm School of Economics
© Elias Olivier and Emanuele De Santis, 2026

Contents

1.	INTRODUCTION	4
1.1.	Background and context	4
1.2.	Wine Labels as point of purchase problem	4
1.3.	Purpose	5
1.4.	Contribution.....	5
2.	THEORY FRAMEWORK AND HYPOTHESIS DEVELOPMENT	6
2.1.	Wine labels as consumer decision cues.....	6
2.2.	Visual semiotics and design codes in wine label communication	7
2.3.	Consumer responses to wine label designs.....	8
2.3.1.	Perceived quality	9
2.3.2.	Purchase intention	10
2.3.3.	Displayed price as contextual cue	11
2.4.	The focal bottle evaluation.....	12
2.5.	Conceptual model (Hypotheses development)	13
2.5.1.	Label style and perceived quality	13
2.5.2.	Label style and purchase intention	14
2.5.3.	Moderating role of price.....	15
2.6.	Summary Table	17
3.	METHOD.....	18
3.1.	Research approach and experimental design	18
3.1.1.	Quantitative approach.....	18
3.1.2.	Experimental logic and causal rationale.....	18
3.1.3.	Design and variables	19
3.2.	Stimulus development	20
3.2.1.	Design rationale and template construction	20
3.2.2.	Typographic and chromatic manipulation.....	21
3.2.3.	Shelf context implementation.....	23
3.2.4.	Pilot purpose, design and procedure.....	24
3.2.5.	Pilot results	25
3.2.6.	Final stimulus selection	26
3.3.	Main study.....	27
3.3.1.	Measures and scales	27
3.3.2.	Procedure.....	28
3.3.3.	Manipulation, instruction, attention checks	29
3.3.4.	Ethical considerations.....	30
3.4.	Data preparation and analysis plan.....	30
3.4.1.	Inclusion and exclusion rules	30
3.4.2.	Scale construction and reliability	30

3.4.3.	Statistical tests	31
3.4.4.	Data screening, analytic sample and randomisation balance	31
4.	ANALYSIS AND RESULTS.....	33
4.1.	Manipulation checks	33
4.2.	Reliability and validity checks.....	34
4.3.	Hypothesis tests.....	36
4.3.1.	H1, Effect of label style on perceived quality	36
4.3.2.	H2, Effect of label style on purchase intention	37
4.3.3.	H3, Moderating role of displayed price on purchase intention	37
4.3.4.	Sensitivity analysis controlling for readability.....	38
5.	DISCUSSION.....	40
5.1.	Findings	40
5.2.	Relation to prior literature.	40
5.3.	Theoretical contribution	42
5.4.	Practical implications.....	42
5.5.	Limitations	43
5.6.	Future research.....	45
6.	CONCLUSION.....	47
	APPENDIX.....	48
	Appendix A. Consideration set decision framework.....	48
	Appendix B. Justification of experimental price levels.....	49
	Appendix C. Robustness checks: ANCOVA with behavioral covariates	50
	Appendix D. AI tools declaration.....	52
	REFERENCES	55

1. Introduction

1.1. Background and context

Wine is a product category marked by strong information asymmetry at the point of purchase, and consumers generally face a large assortment of bottles in stores that differ in ways that cannot be evaluated before consumption. Wine selection is also frequently described as challenging or intimidating, especially for non-expert consumers (Barber et al., 2006). Retail and shelf evaluation of wine remains an important area for study because consumers often make wine decisions without tasting the product first (Lockshin & Corsi, 2012).

Before purchase, consumers rely on extrinsic cues such as price, packaging, and label design to form expectations about the product. Dodd et al. (2005) argue that consumers' knowledge and experience influence which information sources they use, but the front label remains a highly visible source at the point of choice across consumer types. Barber et al. (2006) also found that front label attributes such as country of origin, vintage, and brand name were important to consumers when evaluating wine. The wine label can therefore be understood as a visual interface between the product and the consumer, translating a limited amount of observable information into expectations of quality and purchase suitability.

1.2. Wine Labels as point of purchase problem

The shelf encounter is important because consumers often evaluate wine with little information beyond the bottle, the label, and the displayed price. This pre-purchase settings matters particularly in Sweden, where retail sales of wine and other beverages above 3.5% are organised through the Systembolaget state monopoly (Alkohollag, 2010:1622). Systembolaget's stated mission is to limit alcohol harm by selling alcohol responsibly without a profit motive or advertising (Systembolaget, 2026), which makes the Swedish retail environment more regulated and less promotional than ordinary commercial wine retail. Existing studies show that wine label design influences perceptions, attention, and choice, but they typically use single stimuli or comparative choices rather than evaluation of one focal bottle inside a shelf scene (Sherman &

Tuten, 2011; Lacoste-Badie & Droulers, 2024). Less is known about how modern and traditional label styles work when the consumer evaluates one focal bottle from the label, the bottle, the shelf setting, and the displayed price. The simulated shelf context matters because it brings the evaluation closer to a Systembolaget point of purchase than label stimuli only, while still keeping the focal label and displayed price under experimental control.

1.3. Purpose

The purpose of this thesis is to examine how one selected modern and one selected traditional wine label stimulus, together with displayed price, shape perceived quality and purchase intention when consumers evaluate one focal bottle in a Systembolaget-like shelf context.

1.4. Contribution

This thesis contributes to wine label and pricing research in three connected ways. First, we examine the price to perceived quality cue inside a simulated Systembolaget-like shelf context, where ordinary promotional cues are less central and displayed price remains one of the visible signals available to consumers. Second, we treat perceived quality and purchase intention as separate outcomes instead of combining them into one general evaluation. This lets us examine whether the same cue affects evaluation and stated purchase likelihood in a similar way. Third, we focus the test down to a single focal bottle evaluation instead of a full shelf choice. This isolates a later stage of single bottle evaluation, after attention has already been directed to the bottle.

2. Theory framework and hypothesis development

2.1. Wine labels as consumer decision cues

Wine is a product category characterised by high complexity and competition. Unlike many consumer goods, wine quality cannot normally be assessed before consumption, making the purchase decision uncertain and comparatively high risk (Sherman & Tuten, 2011). This uncertainty is made stronger by the structure of the wine market, since wine brands tend to hold small individual market shares, brand advertising is comparatively limited, and the product is consumed in its commercial packaging (making the label visible in social settings) (Sherman & Tuten, 2011). Consumers frequently report feeling intimidated by the wine selection process (Barber et al., 2006) which reinforces how wine purchasing decisions are made under conditions of uncertainty.

Zeithaml (1988) differentiates between intrinsic cues, which are associated with the physical product itself, from extrinsic cues such as price, brand name, and other product signals outside of the physical product. In the wine context for instance, grape type and region provide product information, while brand name, displayed price, and visual design work as available external cues at the shelf. Among available strategies for reducing uncertainty, the label plays a special role as an accessible and impersonal cue at the point of purchase, and consumers have been shown to seek out wine label information while browsing retail aisles (Sherman & Tuten, 2011). Dodd et al. (2005) suggested a buying decision model in which consumer knowledge and experience shape the use of impersonal, personal, and self-information. In their results, higher (subjective) knowledge is associated with greater reliance on impersonal sources and the self, and with lower reliance on personal sources. We therefore treat the wine label as one important impersonal cue at the point of purchase, not as the only or primary source across all wine consumers and situations. The visual design of the label, including typography, colour, layout, imagery, and overall composition, works not only as decoration but as a communication system that generates holistic impressions. Consumers may use these elements to understand meanings like quality, brand character, modernity, or traditionalism when comparing alternatives. Wine label designs have been categorised into main genres such as traditional designs, contemporary

designs, and novelty designs (Sherman & Tuten, 2011). Taken together, this suggests that wine label design style is an important extrinsic cue system in the consumer decision process, but what remains less clear is how specific visual details can come to communicate meanings such as tradition or modernity. The next section therefore introduces visual semiotics as a framework for understanding this meaning function wine labels produce.

2.2. Visual semiotics and design codes in wine label communication

To understand how wine label design elements carry meaning beyond their aesthetics, we draw on visual semiotics. Visual semiotic research considers elements on product packaging as signs whose material forms, so called signifiers, become associated with concepts and meanings, or signifieds. Packaging design can then be understood as a combination of visual and verbal codes that communicate meanings about the product and its positioning (Celhay & Remaud, 2018; Öztürk & Ertamay, 2019). Applied to wine labels, semiotic analysis has found several categories of visual codes, including layout and composition, colour, illustration, and typography, each of which contributes to the overall meaning a label communicates (Celhay & Remaud, 2018; Öztürk & Ertamay, 2019). These codes operate with structure, as Celhay and Remaud (2018) mapped binaries within each code (such as centred versus non-centred layout, presence versus absence of gold, and complex versus simple holistic design). In an empirical test, they showed that semiotic analysis could anticipate idea associations consumers formed.

Colour is also one of the visual codes with documented semiotic associations in label design for wine. In these mappings, white backgrounds are associated with quality, premium cues, and, in Öztürk and Ertamay's (2019) analysis, purity, while yellowed, ochre, or pale-yellow backgrounds are associated with tradition, seniority, and rustic meanings depending on the study (Celhay & Remaud, 2018; Öztürk & Ertamay, 2019). This chromatic part is relevant to our experiment because the traditional and modern label templates differ not only in typeface but also in background colour.

Beyond colour, typography is important for our study because it is one of the two key visual dimensions through which the traditional and modern stimuli differ. Celhay and Remaud (2018) also treat typography as carrying meaning in broader semiotic systems.

A consistent difference across semiotic label codes is that serif and script typefaces are more conventionally associated with tradition and seniority, while non serif typefaces are linked to more modern impressions (Öztürk & Ertamay, 2019). At the same time, the relationship between modernity and specific type families is not necessarily strict. Cellhay and Remaud's (2018) analysis of Bordeaux wine labels shows that a label can be interpreted as premium, elegant, and modern even when it uses an uppercase non script serif typeface, indicating that modernity is not produced by one rule of typography alone and can instead come from a wider label configuration.

Typography and colour therefore serve as the key semiotic codes in our manipulation. Because these codes work in combination, we operationalise "modern" and "traditional" as holistic perceptions verified through a pilot test (Section 3.2) rather than assumed from individual design features.

2.3. Consumer responses to wine label designs

With the semiotic framework in place, the question is how these communicated meanings translate into consumer responses. In our thesis, the two focal responses are perceived quality and purchase intention, although the second construct must be interpreted carefully. Perceived quality captures respondents' evaluation of the wine before consumption. Purchase intention is kept as the construct label because it is widely used in wine label and marketing research, but our item asks how likely respondents would be to purchase the wine if they saw it at Systembolaget. Strictly, this is closer to stated purchase likelihood or behavioural expectation than to behavioural intention in the stronger sense of a plan or commitment. We therefore use purchase intention as the thesis label, while interpreting the measure as stated purchase likelihood rather than actual behaviour, willingness, or clear behavioural intention. In the following sections we develop the theoretical basis for each construct and for the role displayed price plays as a cue in this context.

2.3.1. Perceived quality

In a wine shelf setting, a quality judgement formed from visible extrinsic cues is effectively the type of evaluation the consumer can make before purchase. Making this pre-purchase judgement the object of this study therefore fits our research question.

Bloch's (1995) model of consumer responses to product form argues that the physical form of a product can generate cognitive responses, such as beliefs about quality, prestige, and value, and affective responses, such as aesthetic pleasure or displeasure. These responses then shape an approach or avoidance behaviour. In this wine context, the label is a central part of the product form that the consumer encounters at the shelf.

Zeithaml's (1988) cue framework also supports this. When intrinsic attributes cannot be evaluated before purchase, consumers make more use of available extrinsic cues. In our thesis, perceived quality is understood as an overall evaluation based on the visible cues available in the Systembolaget-like shelf setting. Steenkamp (1990) refines this further by distinguishing between anticipated fitness for consumption, formed at the point of purchase, and experienced fitness for consumption, assessed upon consumption. Since our thesis examines shelf evaluation rather than tasting, we study the first of the two, but for readability we use the term perceived quality throughout the text. Perceived quality, as used here, is the consumer's summary evaluation of wine quality based on the extrinsic cues available at the shelf (Zeithaml, 1988). Perceived quality may also be shaped by holistic visual impressions that are not directly connected to explicit beliefs about one specific wine label feature. Research on holistic package design supports this view, as Orth and Malkewitz (2008) show that different overall package designs styles are associated with different brand impressions and evaluations.

In our experiment, the two label styles show identical text, so any observed difference in perceived quality between label conditions is expected to reflect the bundle of visual design and the displayed price cue instead of differences in information printed on the label. We do not apply Orth and Malkewitz's specific style categories to our stimuli, but their broader finding that design style relates to consumer impressions supports our use of perceived quality as a primary outcome.

Perceived quality and purchase intention are handled as separate response variables in this thesis because they capture different parts of the consumer response chain.

Perceived quality measures an evaluation of the product, whereas our purchase intention item measures stated purchase likelihood in the specific Systembolaget scenario. Spears and Singh (2004) distinguish brand attitude from purchase intention, and the behavioural intention/expectation difference also suggests that evaluative judgements and expected behaviour should not be collapsed into one general response (Warshaw & Davis, 1985; Sheeran, 2002). We therefore use this distinction as conceptual support for keeping perceived quality and purchase intention as separate outcomes, not as direct evidence that the two are empirically separate in our data.

2.3.2. Purchase intention

Purchase intention is widely used in wine marketing and label research to indicate whether consumers see themselves as likely to buy a specific wine product.

Conceptually, however, purchase intention, willingness, and behavioural expectation are not the same. Behavioural intention refers more strictly to a person's plan or motivational commitment to do a behaviour. Behavioural expectation refers to the person's own estimation of how likely it is that the behaviour will actually occur once real constraints and circumstances are considered (Warshaw & Davis, 1985; Sheeran, 2002). Willingness is also not the best term for our construct, because it can imply readiness to do something effortful or undesirable in order to reach another goal. This distinction matters because our measure asks respondents: "If you saw this wine at Systembolaget, how likely would you be to purchase it?" The item therefore captures stated purchase likelihood in a concrete retail scenario and does not ask whether respondents have formed a plan to buy the wine, whether they would make a special effort to buy it, or whether they would actually spend money on it in a real store.

For consistency with prior label research, we retain the term purchase intention throughout the thesis. However, the results should be interpreted as stated purchase likelihood or behavioural expectation. This means that the findings cannot be read as evidence of actual purchase behaviour or as evidence of a firm behavioural

commitment. Prior research often assumes a positive link between perceived quality and purchase intention, and we deem this assumption theoretically reasonable. A more favourable product evaluation should generally make purchase more likely, and Bloch's (1995) product form model suggests that positive cognitive and affective responses to design can lead to approach behaviour. The Theory of Planned Behaviour similarly places evaluative attitudes before intention (Ajzen, 1991). Still, the relationship should not be treated as automatic, since a respondent may evaluate a wine as high quality without expecting to purchase it because the occasion, budget, alternatives, personal taste, or consumption situation do not fit. We therefore analyse perceived quality and purchase intention as separate dependent variables rather than modelling perceived quality as a mediator. This is both a conceptual and design choice, since the study was designed to test direct effects of label style and displayed price on each outcome, and both outcomes were measured after the same stimulus exposure rather than in a sequence over time that would support a strong mediation claim. A possible perceived-quality-to-purchase-likelihood link is therefore acknowledged theoretically, but a formal mediation model falls outside the scope of our study.

2.3.3. Displayed price as contextual cue

Outside of label design, displayed price is the other directly visible cue consumers encounter at the wine shelf. Price represents a cost but also a signal shaping how consumers interpret the other cues around it. This double role is described in the price/quality/value framework proposed by Zeithaml (1988) and tested by Dodds et al. (1991), in which price had a positive effect on perceived quality but a negative effect on perceived value and willingness to buy. Another relevant factor influencing product evaluations in Dodds et al. (1991) is store information. This is relevant here because Systembolaget is part of the evaluation context, even though its regulated monopoly role is not identical to the store manipulation used in that experiment.

In wine, price is often an influential cue for less expert consumers. Lockshin and Rhodus (1993) found that consumers responded positively to higher price even when tasting was possible, and that price was a significant predictor in their model. In a monopoly shelf context without any possibility for tasting, where consumers evaluate a

wine selection almost exclusively with visible price and package cues, the role of displayed price may be even more important. Spence (2024) in a similar way reviews evidence showing that price information can change wine evaluations through expectation effects (especially for low to medium priced wines). The strength of the price link to quality is also not automatic, however. Völckner and Hofmann's (2007) analysis showed that the price to perceived quality relationship varies by product category, study design, respondent familiarity, and country sample, meaning that the price should not only be treated as a direct cue, but also as something that may change how other visible cues are interpreted, and we will return to this in Section 2.5 (where displayed price is tested as a possible moderator of the label style effect on purchase intention).

2.4. The focal bottle evaluation

The constructs developed above describe the main cues wine consumers use and the responses those cues produce, but they do not specify where in the choice process these mechanisms work. Because our study tests label and price effects at a particular stage rather than across the full Systembolaget choice journey, we now describe that stage more explicitly.

Consideration set theory is useful for modelling this stage, because consumers in large assortments such as Sweden's state monopoly are unlikely to evaluate every available option in detail and instead tend to narrow the choice down to a smaller set before making a final decision (Shocker et al., 1991). The present experiment does not observe this narrowing directly. Instead, it captures a specific moment of inspection, where respondents are placed in a Systembolaget like shopping scenario and asked to evaluate one displayed bottle. The study can therefore comment on how label style and displayed price shape evaluation once attention has already been directed at a specific bottle, but it cannot show how labels first attract attention, how wines enter the consideration set, or whether the Systembolaget context itself causes any observed effects. Two illustrations of the consideration set logic behind the experiment are provided in Appendix A .

2.5. Conceptual model (Hypotheses development)

We now develop the expected effects of label style and displayed price on perceived quality and purchase intention. Drawing on these theoretical foundations, we suggest a model in which label style may affect perceived quality and purchase intention, but where these effects may be different in outcome and may depend on displayed price. Our model uses visual semiotics (Celhay & Remaud, 2018), product form theory (Bloch, 1995), holistic design perception effects (Orth & Malkewitz, 2008), and the price cue literature (Dodds et al., 1991). The central idea from this literature review is not that modern labels are automatically better than traditional labels but that modern and traditional labels may activate different meanings, and that these meanings may matter at different levels for quality judgment and purchase intention judgments when price is visible.

2.5.1. Label style and perceived quality

The expected effect of label style on perceived quality is ambiguous from a theory point of view because modern and traditional wine labels can both carry favourable meanings, but through different mechanisms. Modern labels can benefit from distinctiveness, novelty, visual simplicity, and moderate category incongruity. Bloch's (1995) product form model suggests that product form can create cognitive and affective responses, and the MAYA logic suggests that designs that are sufficiently "advanced" to be interesting but still "acceptable" within the category can generate favourable evaluations (Hekkert et al., 2003). Celhay et al. (2017) apply similar reasoning to wine labels, where appropriate deviation from category conventions may help a label stand out without becoming illegible as wine.

In our setting, the modern label is therefore expected to benefit if respondents experience the cleaner and more contemporary design as distinctive while still credible as a Systembolaget wine. Recent research in alcohol labels also supports this directional expectation. Lacoste-Badie and Droulers (2024) found that young adults looked more quickly at modern alcohol labels, revisited, looked at them for longer, and selected beer with modern labels more often than with traditional labels. Their study differs from ours because it examined attention and comparative selection for beer/alcohol-label stimuli

rather than perceived quality in a focal wine-bottle evaluation, but it provides a relevant mechanism since modern label cues may win earlier in the evaluation process by attracting attention and supporting selection among young consumers.

At the same time, traditional labels may also generate favourable quality evaluations. Prior wine research connects traditional visual codes to authenticity, heritage, category membership, and associations of higher status. Boudreaux and Palmer (2007) found traditional labels linked to higher perceived value and upper-class brand personality, while Sherman and Tuten (2011) describe traditional wine imagery such as coats of arms and vineyards as meaningful label cues. Because wine quality cannot be verified before consumption, familiar category cues may also reduce perceived risk, especially for consumers who want reassurance that the bottle fits the category. Celhay and Passebois (2011) and Celhay and Trinquescoste (2015) in addition suggest that typicality can become more attractive when choice feels risky.

H1 is therefore directional, and the argument is not that modern labels are universally better than traditional labels, but our narrower prediction is that, in the focal stage studied here, moderate modern distinctiveness may generate a more favourable perceived quality evaluation than the more familiar traditional bundle. However, a traditional advantage would also be theoretically interpretable if the authenticity, familiarity, category fit, or risk meanings were to win the distinctiveness advantage.

H1. In a Systembolaget-like shelf context, the modern label will generate higher perceived quality than the traditional label.

2.5.2. Label style and purchase intention

H2 applies the same directional logic to purchase intention, interpreted here as stated purchase likelihood. If the modern label creates a more favourable response through distinctiveness, visual clarity, and contemporary design associations, this may also increase respondents' expected likelihood of purchasing a bottle. Product form theory supports this general connection because positive cognitive and affective responses to design can lead to approach behaviour (Bloch, 1995). The link is not assumed to be

automatic, however, because perceived quality and stated purchase likelihood are conceptually different outcomes.

Directional support for a modern label advantage also comes from Lacoste-Badie and Droulers (2024), who found that modern alcohol labels attracted visual attention more strongly and that beer bottles with modern labels were selected more often than those with traditional labels among young adults. Their study does not directly test our dependent variable, because attention and comparative selection are not the same as stated purchase likelihood in a focal wine-bottle evaluation. They suggest however that modern label cues in alcohol can produce favourable consumer responses under some conditions. The traditional label counterargument remains important, since traditional label could increase purchase likelihood if that makes the wine feel safer, more authentic, more category-congruent, or better matched to a higher price. H2 is therefore a directional hypothesis based on the mechanisms expected to be stronger in our focal context and not a claim that traditional labels should be ineffective on a general level.

H2. In a Systembolaget-like shelf context, the modern label will generate higher purchase intention than the traditional label.

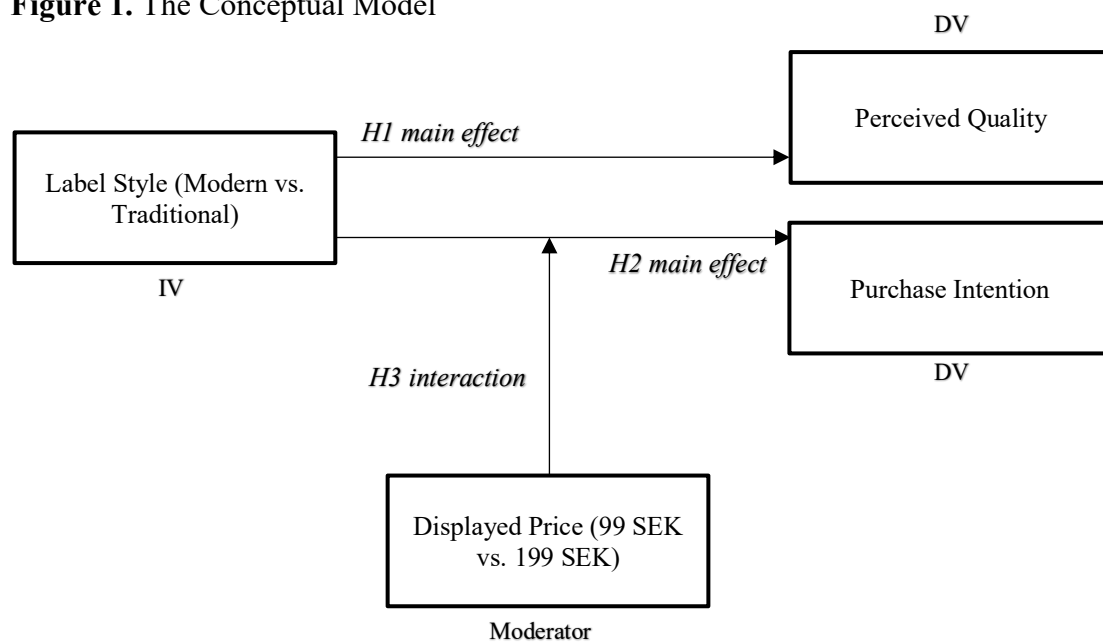
2.5.3. Moderating role of price

We argue that displayed price may change how a label style is read and acted on, calling this mechanism cue fit. Traditional label codes are often connected to traditionalism and familiar wine images, while traditional layouts have been associated with higher perceived value and upper-class brand personality (Boudreaux & Palmer, 2007; Sherman & Tuten, 2011), and these meanings may fit better when the bottle is shown at a higher price. At the lower display price, this premium fit effect should be weaker, and the modern label may benefit more from being clear, simple, and contemporary. At the higher price, the traditional label may become relatively more persuasive because the higher price and the more classic design point in a similar cue direction. We do not predict that the traditional label outperforms the modern label at the higher price necessarily, but do predict that the modern label's purchase intention advantage over the traditional label is smaller at 199 SEK than at 99 SEK.

H3 uses purchase intention as the dependent variable for two reasons. First, the price main effect on perceived quality is already observable in the H1s ANOVA model, and second, cue fit is more about how two visible cues work together at the moment of choice (and both Bloch (1995) and Spears and Singh (2004) treat the intention to buy as separate from the product evaluation that precedes it). Dodds et al. (1991) also found that in the same experiment, price raised perceived quality but lowered willingness to buy, showing that price effects do not necessarily have to move both outcomes in the same way.

H3. In a Systembolaget-like focal bottle shelf evaluation, displayed price moderates the effect of label style on purchase intention, such that the purchase intention advantage of the modern label over the traditional label is smaller at 199 SEK than at 99 SEK.

Figure 1. The Conceptual Model



2.6. Summary Table

Table 1. Summary of hypotheses and variables.

	Hypotheses	IV	DV
H1	In a Systembolaget-like shelf context, the modern label will generate higher perceived quality than the traditional label.	Label style	Perceived quality
H2	In a Systembolaget-like shelf context, the modern label will generate higher purchase intention than the traditional label.	Label style	Purchase intention
H3	In a Systembolaget-like focal bottle shelf evaluation, displayed price moderates the effect of label style on purchase intention, such that the purchase intention advantage of the modern label over the traditional label is smaller at 199 SEK than at 99 SEK.	Label style x displayed price	Purchase intention

Note. IV = independent variable; DV = dependent variable, H3 tests the label style x displayed price interaction on purchase intention. Label style is modern vs. traditional. Displayed price is 99 SEK vs. 199 SEK.

3. Method

3.1. Research approach and experimental design

The conceptual model we have developed in Section 2 specifies two manipulation cues (label style and displayed price) and two response variables (perceived quality and purchase intention). The rest of this section describes the experimental design used to test the three hypotheses derived from our model.

3.1.1. Quantitative approach

We adopted a quantitative experiment approach because the research question deals with the causality effects of label style on consumer responses, which is a question best answerable with controlled manipulation and comparison (Söderlund, 2018). Our experiment satisfies the main conditions for causal study in a controlled experiment since the treatment precedes the response, alternative explanations were limited through random assignment and stimulus control, and covariation is tested statistically (Söderlund, 2018).

3.1.2. Experimental logic and causal rationale

An experimental approach allows us to manipulate the independent variable directly and to keep the remaining conditions of our study under control, which is what makes it possible to isolate the effect of wine label style and limit the influence of potential confounding variables. We implement this logic in different design decisions for this study.

First, all six pilot stimuli (and the two main stimuli derived from those) show identical text and differ only in typeface and background colour combinations that make up the label style manipulation, so that the rest of the label is visually constant. Additionally, the scenario text alongside the stimulus is held identical across all conditions so that any difference in perceived quality or purchase intention observed cannot be linked back to differences in what respondents read alongside the image. Lastly, random assignment was implemented through the internal randomiser in Qualtrics (the online survey platform used to administer the experiment) which distributed across conditions. We

acknowledge that the manipulation does not isolate the separate effects of typeface, colour, or readability, since we test label style as a holistic bundle of visual cues, but still support a causal test of the bundle as a whole.

3.1.3. Design and variables

We conducted a 2x2 between-subjects fully online Qualtrics experiment. Label style had two levels, modern and traditional. Displayed price also had two levels, 99 SEK and 199 SEK. Each participant was randomly assigned to one of four conditions and evaluated only one bottle.

The 99 SEK and 199 SEK values were chosen from Systembolaget's public launch plans reported in Appendix B, where the generic high volume white and red wine category is divided into recurring price ranges from 80-99 SEK at the low end to up to 199 SEK at the higher end. Selecting values from the two extremes within the same generic category kept both displayed prices realistic for a Systembolaget shelf while still providing a clear price difference between generic conditions. Because the scenario told respondents beforehand that they had decided on a budget matching the displayed price, the price manipulation should be interpreted as displayed price inside a budget-consistent focal evaluation instead of a more general test of price sensitivity.

The dependent variables were perceived quality, measured as a three-item index, and purchase intention, measured with a single item. Both were measured on seven-point Likert scales; exact item wording is reported in Table 5. Purchase intention should be interpreted as stated purchase likelihood / behavioural expectation instead of observed behaviour, willingness, strict behavioural intention, or actual purchase behaviour.

Making label style a between-subjects factor was justified by concerns we had about demand effects in wine research. With the same participant evaluating both a modern and a traditional identical label, the very structure of the questions could make the experiment's goal obvious. Participants who notice the modern-versus-traditional study may figure out the intention of our work and change their answers in response to a perceived purpose (Söderlund, 2018). A fully between-subjects design responds to this

risk because each participant only sees one bottle and has no direct or obvious way to identify a modern-versus-traditional comparison that the study is built for.

Displayed price was kept between-subjects for the same reason (showing the same participant both a 99 SEK and 199 SEK versions of the same wine would create a direct price comparison that our study is not interested in testing). Combining the two between-subjects factors gives a clean 2x2 design in which each cell contains independent participants. Our design has lower statistical power than a within-subjects alternative for a given sample size, but it gives us stronger protection against demand effects. Figure 2 gives a simple overview of the four conditions created combining label style and displayed price.

Figure 2. 2x2 between-subjects experiment design overview

		Displayed Price	
		99 SEK	199 SEK
Label Style	Traditional	Condition 1	Condition 2
	Modern	Condition 3	Condition 4

3.2. Stimulus development

3.2.1. Design rationale and template construction

To make modern and traditional perceptions possible to test in the pilot study, six fictional pilot labels stimuli were designed, with three labels in each category. The stimuli were prepared using Canva, a browser graphic design platform. Each category uses a consistent label template. The two templates differ in two ways motivated by the semiotic literature explored, and that is specifically the typeface and the background colour of the label. Background colour was held constant within each category but changed between the traditional and modern labels.

The traditional labels use a warm cream background, consistent with semiotic associations between yellowed, ochre, or yellow backgrounds and tradition, seniority, or rustic meanings in prior wine mappings (Celhay & Remaud, 2018; Öztürk & Ertamay, 2019). The modern template uses a clean white background, in line with associations between white, quality, premium cues and purity (Öztürk & Ertamay, 2019). All other design elements within each template, including layout, text content, and smaller text elements, are held constant. This means that the manipulation should be understood as a holistic label manipulation. We do not isolate the separate causal effect of typeface or background colour alone, because these elements work together to create the modern or classic impression. All label elements were held constant within each category to control for confounds outside of typeface and background colour.

3.2.2. Typographic and chromatic manipulation

Within each label category, the three pilot variants differed only in the typeface used for the cuvée name. The traditional variants used Symphony, Libre Baskerville, and Snell Roundhand on the warm cream template. The modern variants used Poppins, Avenir, and League Spartan on the clean white template. The cuvée name “Cuvée Umbra” was set in bold in all variants, keeping typography weight constant so that differences within each category mainly reflected typeface form instead of differences in emphasis.

Figure 3. The six pilot stimuli



3.2.3. Shelf context implementation

In our main experiment the focal bottle was shown inside a shelf image meant to resemble a Systembolaget wine section, rather than on its own. This kept the wine in a retail setting while the focal label and displayed price stayed as the main cues (surrounding bottles were visible but made less visually salient, so they would not compete with the focal bottle or trigger brand associations).

Each respondent saw only one focal bottle, based on the label style and price condition Qualtrics assigned them. The image was shown alongside the following scenario text: "You are shopping at Systembolaget and have decided on a budget of 99/199 SEK. You pick up the bottle shown above. Please answer the following questions based on your impression of this wine". This setup fits our focus on focal bottle evaluation rather than full shelf choice. It keeps some of the Systembolaget context while still giving the experiment enough control, since a fully detailed shelf with many readable labels could have introduced visual competition the study was not designed to test. The scenario text and shelf background were identical across all four conditions, with the only differences being the focal label style and the displayed price, which appeared both in the scenario wording and as a price tag embedded at the bottom of the stimulus image.

Figure 4. Focused bottle shelf context, aisle desaturated. The displayed price appeared both in the scenario wording and as a price tag embedded at the bottom of the stimulus image. Both labels are shown here at 99 SEK; the 199 SEK conditions followed the same layout with only the price tag updated.



3.2.4. Pilot purpose, design and procedure

We conducted a pilot test prior to our main experiment to verify that the six label stimuli were perceived as intended, ranging from modern to classic, and to measure any unintended effects from confounding variables that could impact the main study's validity. Establishing successful manipulation before hypothesis testing was an important methodological step for the label stimuli, because it functioned as our evidence that differences in dependent variables could be linked to the semiotic manipulation rather than to other stimulus variation outside of our control (Perdue & Summers, 1986).

The pilot test was conducted online through Qualtrics. Respondents were shown all six label stimuli one at a time in fully randomised order, so that modern and traditional variants were mixed. A within-subjects design was used to maximise statistical power given the small pilot sample, while full order randomisation was applied to reduce potential "carryover" effects for the labels being shown.

For each label, respondents answered three sets of questions. The first set worked as manipulation check and consisted of four seven-point Likert-scale items: "This label looks traditional", "This label looks modern", "This label looks classic", and "This label looks contemporary". The second set addressed stimulus plausibility and potential confounds, and asked three questions measured on 7-point Likert scales: "This wine could realistically be sold at Systembolaget", "This label fits the type of wine usually seen at Systembolaget", and "The information on this wine label is easy to read". The third part was an open question asking what specifically made the label look modern or traditional, following the free association previously used in wine research to identify visual cues which respondents rely on when interpreting labels (Celhay & Remaud, 2018).

After evaluating all six labels, respondents completed a brief demographics section covering age, gender, frequency of wine purchase at Systembolaget, and frequency of wine consumption. All responses were collected anonymously, and respondents confirmed that they were at least 20 years of age before proceeding. Success in our manipulation was defined as traditional labels scoring better on the traditional and

classic items and lower on the modern and contemporary items than the modern labels, with the opposite true for the modern category. The confound checks were used to verify no differences in readability on a systematic level.

3.2.5. Pilot results

Of 44 respondents who started the pilot survey, 32 completed all items and formed our final sample. The pilot sample was predominantly young as 47% were aged 20-23, 22% aged 24-27, 6% aged 28-30, and 25% above 30. The sample was 59% male and 41% female. Most respondents purchased wine at Systembolaget a few times a year (41%) or a few times a month (28%). 13% reported never purchasing there, while the remaining 18% reported other purchase frequencies.

Table 2. Pilot manipulation check: mean scores (SD) per label across four items (1-7 Likert scale)

Label	Traditional	Modern	Classic	Contemporary	Outcome
<i>Traditional</i>					
Symphony	5.25 (1.46)	2.91 (1.33)	5.12 (1.54)	3.34 (1.58)	Pass *
Libre Baskerville	3.78 (1.60)	4.34 (1.64)	4.16 (1.46)	4.59 (1.54)	Ambiguous
Snell Roundhand	4.72 (1.75)	3.28 (1.57)	4.81 (1.75)	3.62 (1.56)	Pass
<i>Modern</i>					
Avenir	2.38 (1.24)	5.62 (1.48)	2.81 (1.53)	5.12 (1.43)	Pass
Poppins	2.47 (1.48)	5.81 (1.33)	2.47 (1.32)	5.12 (1.60)	Pass *
League Spartan	2.16 (0.92)	5.38 (1.60)	2.22 (1.10)	5.19 (1.38)	Pass

Note. $N = 32$. Values are means with their standard deviations in parentheses. The outcomes show whether each label followed intended modern/traditional across all four manipulation items. Asterisks show the stimuli was kept for the main experiment.

At the category level, the pattern we intended appeared clear across the manipulation items in Table 2. Five of the six labels performed as intended. Libre Baskerville was the exception, since it received more mixed answers and was therefore treated as ambiguous rather than as a strong traditional or heritage stimulus.

Table 3. Pilot confound check means by label category (1-7 Likert scale)

Confound item	Traditional M	Modern M	Mean Difference
Systembolaget plausibility	5.59	5.05	Small
Systembolaget fit	5.17	4.49	Small
Readability*	5.27	6.21	Meaningful

Note. $N = 32$. Values are category means for the three pilot labels in each category. Items were measured on 1-7 Likert scales so the scale midpoint is 4. The table is **descriptive**.

The labels were generally seen as credible Systembolaget products, since plausibility and fit were above the scale midpoint. The main issue was readability, where the modern labels were easier to read than the traditional labels as per Table 3. Open answers presented in Table 4 also showed that respondents mainly used typography and background colour when judging whether a label looked modern or classic, which supports the intended semiotic manipulation.

Table 4. Dominant three cues mentioned in open pilot responses, by label category

Traditional labels	Modern labels
Cursive or script typography	Bold sans-serif typography
Warm cream or yellow background	Clean white background
Ornamental or decorative letterforms	Minimalist overall design

Note. $N = 32$. Entries summarise the dominant visual cues mentioned in our open responses. Frequencies are not reported and the table is only descriptive.

3.2.6. Final stimulus selection

Based on the pilot results, one label from each category was selected for the main experiment, as indicated in Table 2. Symphony was chosen as the traditional stimulus because it produced the clearest traditional and classic score and the lowest modern and contemporary value among the traditional candidates. Poppins likewise was selected as the modern stimulus because it had the strongest modern and contemporary profile between the modern candidates. Libre Baskerville was excluded because it was read as ambiguous from the four manipulation answers.

Our final pair maximised the separation between the two intended label styles while still showing sufficient Systembolaget plausibility. Our choice also meant, however, that readability was not perfectly matched between the selected styles, so readability was remeasured in the main study and treated as an important sensitivity check for evaluations. The main study therefore tests one selected modern label stimulus and one selected traditional label stimulus. The pilot results should be interpreted as evidence for our specific operationalisation of modern versus traditional label style, not as evidence for all possible modern and traditional wine labels.

3.3. Main study

3.3.1. Measures and scales

Table 5 presents all measures from the main survey, organised by role as dependent variables, manipulation checks, confound and credibility items, exclusion checks, and finally demographics. Perceived quality was measured as the average of three items, consistent with Zeithaml's (1988) treatment of perceived quality as summary evaluation from extrinsic cues. Full item wording, roles and scales are reported in Table 5.

Purchase intention was measured with one item: "If you saw this wine at Systembolaget, how likely would you be to purchase it?" The construct is retained as purchase intention for consistency with prior label research, but our item is interpreted as stated purchase likelihood / behavioural expectation rather than as willingness, strict behavioural intention, or actual purchase behaviour.

Table 5. Survey measures: constructs, roles, item wording, and scales

Construct	Role	Item wording	Scale
Perceived quality	DV	"This wine appears to be of high quality" "This wine seems like a good product" "This wine gives a premium impression"	7-point Likert
Purchase intention	DV	"If you saw this wine at Systembolaget, how likely would you be to purchase it?"	7-point Likert
Label style perception	Manipulation check	"This label looks traditional" "This label looks modern" "This label looks classic" "This label looks contemporary"	7-point Likert
Price perception	Manipulation check	"The displayed price seems realistic for a wine at Systembolaget" "The displayed price seems expensive for a wine at Systembolaget"	7-point Likert
Stimulus credibility and readability	Confound check	"This wine could realistically be sold at Systembolaget" "This label fits the type of wine usually seen at Systembolaget" "The information on this label is easy to read"	7-point Likert
Instruction check	Exclusion	"In this study, you were asked to evaluate: (a) Wine Labels (b) Beer Cans (c) Pizza Packaging (d) Coffee Packaging"	Multiple choice
Attention check	Exclusion	"To confirm you are reading carefully, please select the number 5 below"	7-point Likert
Age	Demographic	Under 20 / 20-23 / 24-27 / 28-30 / Above 30	Category
Gender	Demographic	Male / Female / Non-binary/third gender / Prefer not to say	4-option
Wine purchase frequency	Demographic	"How often do you purchase wine at Systembolaget?"	5-point ordinal

Wine consumption frequency	Demographic	“How often do you drink wine?”	5-point ordinal
Wine knowledge	Demographic	“How would you rate your knowledge of wine?”	7-point Likert

Note. DV = dependent variable. The instruction and attention checks were used for exclusion decisions. Exact item wording is shown in quotation marks.

3.3.2. Procedure

Figure 5. Survey procedure and randomisation into experimental conditions

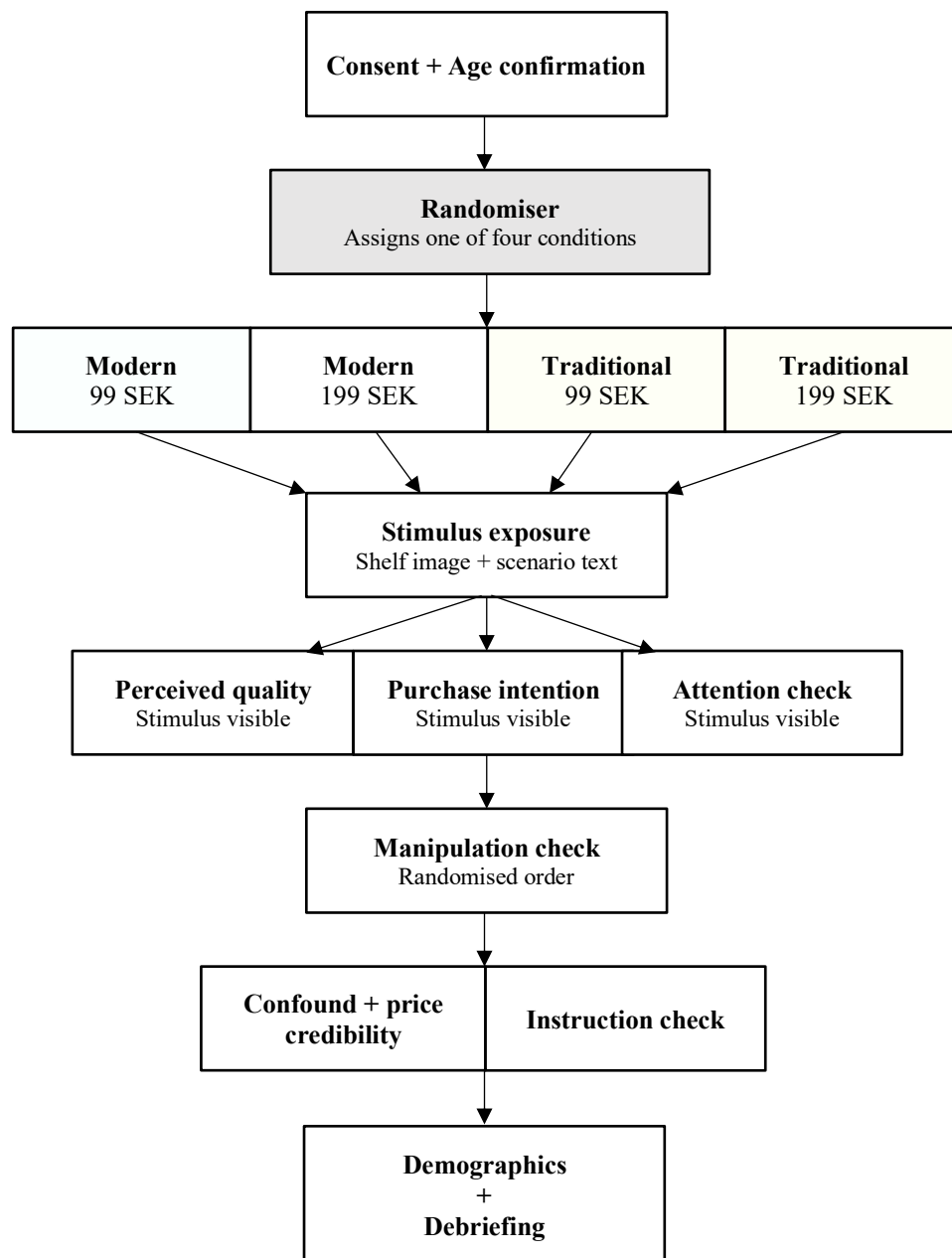


Figure 5 presents the full survey procedure followed. Recruitment for the main study was kept separate from the pilot study (negatively impacting the final number of our collected answers), so that no respondent participated in both phases. This was done to avoid carryover effects, since pilot respondents had already been exposed to the modern and traditional manipulation and would have likely understood our experimental purpose, biasing their responses in the main study. Our recruitment also specifically targeted respondents who were familiar with Systembolaget and had potentially purchased wine there, so that the simulated shelf scenario would be at least interpretable. After consent and age confirmation, Qualtrics' randomiser assigned each respondent to one of our four conditions before any stimulus was shown. During the perceived quality and purchase intention items, the stimulus image remained visible on the same page. This was a deliberate choice, as removing the image after exposure would force respondents to evaluate the wine from memory instead of direct visuals, introducing a recall confound that is not present in the actual shelf encounter our experiment studies. The survey closed with a short debriefing informing respondents that the label stimuli were fictional and not associated with any real wine producer or region. The exclusion criteria applied to the resulting data are set out in Section 3.4.1.

3.3.3. Manipulation, instruction, attention checks

Because the main study used a 2x2 between-subjects design, the manipulation checks were evaluated using 2x2 ANOVA models rather than only independent-samples t-tests. For the label manipulation, the four label-perception items (“traditional”, “modern”, “classic”, and “contemporary”) were analysed with label style and displayed price as between-subjects factors. A successful label manipulation should show the intended main effect of label style while showing no meaningful spillover from displayed price and no label style x displayed price interaction. For the price manipulation, the price-perception items were analysed using the same 2x2 logic, with the “expensive” item treated as the central manipulation check and the “realistic” and “fits the wine” items treated as price credibility checks.

The manipulation checks were shown after the dependent-variable measures so that the modern/traditional contrast and the price perception questions would not reveal the

purpose of our experiment before respondents answered the perceived quality and purchase intention items. The instruction validity check and attention check were included in separate blocks to identify respondents who misunderstood the task or did not read the survey items properly. Responses failing these checks were excluded because they could not provide valid evidence about the dependent variables (exclusion rules applied to these checks are described in Section 3.4.1.).

3.3.4. Ethical considerations

All participants provided informed consent before beginning the survey by confirming, on the opening Qualtrics page, that they were at least 20 years of age and agreed to participate voluntarily, and respondents could also withdraw at any time without consequences. Our study did not involve deception in a strict sense, as participants were asked to evaluate wine labels in a hypothetical purchase scenario. The participants were not aware that our labels were fictional until the debriefing statement at the end of the survey, however. The study complied with GDPR requirements and institutional ethical guidelines (no personal data was in any way collected, processed nor stored), and a separate Appendix D will provide a declaration of AI tools used in our process.

3.4. Data preparation and analysis plan

3.4.1. Inclusion and exclusion rules

We specified a list of exclusion criteria before we collected data, and responses were excluded if a participant was under 20 years of age, did not complete the full survey, failed the instruction validity check by selecting a wrong response to the instruction question, or failed our included attention check. We did not specify a time threshold, and response times were inspected visually but resulted in no additional exclusions.

3.4.2. Scale construction and reliability

The three perceived quality items were averaged into a single score per participant, and internal consistency for this score was assessed using coefficient alpha, with .70 used as a conventional benchmark instead of an absolute criterion for acceptable reliability (Greco et al., 2018). We measured purchase intention with a single item, so no reliability estimate was calculated for this variable.

3.4.3. Statistical tests

We analysed the final data using 2x2 between-subjects ANOVA models with label style (modern vs traditional) and displayed price (99 SEK vs 199 SEK) as between-subjects factors. The manipulation checks and credibility checks were analysed with the same factorial structure so that we could test not only whether the intended manipulation worked but also whether one factor spilled over into the manipulation check for the other factor. H1 and H2 were tested with two separate 2x2 ANOVA models, using perceived quality and purchase intention as dependent variables. H1 was evaluated through the main effect of label style on perceived quality, H2 through the main effect of label style on purchase intention, and H3 through the label style x displayed price interaction in the purchase intention model. Type III sums of squares were used because the cell sizes were similar but not identical and our effect sizes are reported as partial eta squared.

3.4.4. Data screening, analytic sample and randomisation balance

A total of 136 respondents completed the main survey. Eleven respondents were excluded before analysis because they failed the attention check, and one of these also failed our instruction validity check. This left a final sample of 125 participants for analysis, with the four experimental cells balanced (with cell sizes between 30 and 32 respondents, as shown in Table 6).

The sample was mostly young, with 60.8% of respondents aged 20 to 23, which reflects the use of our social media and personal networks for recruitment. Respondents also reported moderate wine knowledge on average, with a mean of 3.70 on the Likert scale. Our balance tests showed no significant differences across conditions for age, gender, wine purchase frequency, wine consumption frequency, or wine knowledge, as shown in Table 7. This suggests random assignment produced comparable groups, but with small category cells the balance tests are descriptive and not strong evidence of fully equal groups. Consistent with the recruitment frame, almost all respondents reported some experience purchasing wine at Systembolaget, with only 10.4% reporting they had never done so.

Table 6. Sample and condition characteristics

Characteristics	Categories	n	%
Condition	Traditional 99 SEK	31	24.8
	Traditional 199 SEK	30	24.0
	Modern 99 SEK	32	25.6
	Modern 199 SEK	32	25.6
Age band	20-23	76	60.8
	24-27	14	11.2
	28-30	7	5.6
	Above 30	28	22.4
Gender	Female	52	41.6
	Male	70	56.0
	Prefer not to say	3	2.4
Wine purchase at Systembolaget	A few times a year	36	28.8
	A few times a month	32	25.6
	About once a month	31	24.8
	Never	13	10.4
	Once a week or more	13	10.4
Wine consumption frequency	A few times a month	44	35.2
	A few times a year	28	22.4
	About once a month	19	15.2
	Once a week or more	27	21.6
	Never	7	5.6
Self-rated wine knowledge		M = 3.70 (SD = 1.54)	

Note. $N = 125$. Percentages are based on the final analytic sample and may not sum up to 100 because of rounding. Wine knowledge was measured on a 1-7 scale.

Table 7. Balance test results across experimental conditions

Variable	Test	Statistic	df	p
Age band	χ^2	5.24	9	0.813
Gender	χ^2	7.46	6	0.281
Wine purchase frequency	χ^2	19.67	12	0.074
Wine consumption frequency	χ^2	15.07	12	0.237
Self-rated wine knowledge	F	1.71	3, 121	0.169

Note. Balance tests compare the four experimental conditions in the $N = 125$. Chi-square tests were used for category variables and one-way ANOVA for wine knowledge. *P*-values are two sided.

4. Analysis and Results

This section reports the manipulation checks, reliability and validity checks, and hypothesis tests. The sample, exclusion decisions, and randomisation balance are reported in Section 3.4.4. The exact test statistics are reported in tables to keep the text readable. All analyses were conducted in R Studio, mainly using the *afex* and *emmeans* packages. Type III sums of squares were used for the ANOVA models, and effect sizes are reported as partial eta squared.

4.1. Manipulation checks

The label manipulation worked as intended, and in the 2x2 ANOVAs, label style had significant main effects on all four perception items. Respondents who saw the traditional label rated it as more traditional and more classic, while respondents who saw the modern label rated it as more modern and more contemporary. Displayed price did not significantly affect any of the four perception items, and none of the label style x displayed price interactions were significant. This supports the interpretation that the label manipulation worked without statistical evidence of meaningful spillover from the price manipulation. The central price manipulation also worked as intended, and respondents in the 199 SEK condition rated the displayed price as more expensive than respondents in the 99 SEK condition, with no significant main effect of label style and no significant label style x displayed price interaction on the expensive-price item. The two credibility items for price add some caution, however. Both prices were generally seen as realistic, but the modern label was rated as marginally more realistic overall, and the 199 SEK price was rated as fitting the wine less well than the 99 SEK price. The price manipulation is therefore treated as successful for perceived expensiveness, while the realism and fit checks are interpreted as credibility caveats instead of failed manipulations.

Table 8. Label manipulation checks from 2x2 ANOVAs

Item	Traditional label M (SD)	Modern label M (SD)	Label style result	Price <i>p</i>	Interaction <i>p</i>
Traditional	4.98 (1.68)	2.66 (1.73)	$F = 57.34, p < 0.001, \eta^2 = 0.322$	0.564	0.864
Modern	3.25 (1.48)	5.86 (1.40)	$F = 101.57, p < 0.001, \eta^2 = 0.456$	0.928	0.452
Classic	5.25 (1.34)	3.14 (1.90)	$F = 50.70, p < 0.001, \eta^2 = 0.295$	0.205	0.601
Contemporary	3.62 (1.31)	5.47 (1.32)	$F = 60.71, p < 0.001, \eta^2 = 0.334$	0.826	0.759

Note. $N = 125$. Values are means and standard deviations on 1–7 scales. Each item was tested with a 2x2 ANOVA using label style and displayed price as between-subjects factors. All *F*-tests in the label style result column have $df = 1, 121$. η^2 = partial eta squared.

Table 9. Price manipulation and credibility checks from 2x2 ANOVAs

Item	99 SEK M (SD)	199 SEK M (SD)	Displayed price result	Label style <i>p</i>	Interaction <i>p</i>
Expensive	2.48 (1.66)	4.02 (1.74)	$F = 25.50, p < 0.001, \eta^2 = 0.174$	0.589	0.390
Realistic	5.63 (1.52)	5.21 (1.61)	$F = 2.35, p = 0.128, \eta^2 = 0.019$	0.053	0.472
Fits the wine	5.44 (1.54)	4.71 (1.67)	$F = 6.59, p = 0.012, \eta^2 = 0.052$	0.094	0.067

Note. $N = 125$. Values are means and standard deviations on 1–7 scales. Each item was tested with a 2x2 ANOVA using label style and displayed price as between-subjects factors. All *F*-tests in the displayed price result column have $df = 1, 121$. η^2 = partial eta squared. The “expensive” item is the central price manipulation check. “Realistic” and “fits the wine” are credibility checks and are interpreted more cautiously.

4.2. Reliability and validity checks

The three perceived quality questions were averaged into one measure and showed high internal consistency, with coefficient alpha = 0.939, well above the conventional 0.70 benchmark discussed in the analysis plan. Purchase intention was measured with one item, so no reliability estimate was calculated for that variable.

The ANOVA checks were acceptable overall. Our Levene's tests did not show unequal variances for perceived quality or purchase intention, and perceived quality residuals followed approximately a normal distribution. Purchase intention showed a small deviation from normality (Shapiro Wilks $p = 0.011$), but ANOVA is robust when cell sizes are balanced and each cell exceeds approximately 30 observations (which our design satisfies). The main remaining stimulus issue was readability. The modern label was rated more readable than the traditional label at both prices, with a larger gap at 99 SEK than at 199 SEK. This pattern is reflected in the significant label style x displayed price interaction shown in Table 11, and because this readability gap could affect interpretation of the label style results, the sensitivity analysis in Section 4.3.4 should be read together with the main H1 and H2 tests.

Table 10. Reliability and ANOVA assumption checks

Check	Test	Result	Interpretation
Perceived quality reliability	Cronbach's alpha	Alpha = 0.939	High internal consistency
Equal variances, perceived quality	Levene's test	$F(3, 121) = 2.06$, $p = 0.109$	No significant violation
Equal variances, purchase intention	Levene's test	$F(3, 121) = 0.55$, $p = 0.647$	No significant violation
Normality, perceived quality residuals	Shapiro Wilk	$W = 0.986$, $p = 0.226$	No significant deviation
Normality, purchase intention residuals	Shapiro Wilk	$W = 0.972$, $p = 0.011$	Small deviation

Note. $N = 125$. Reliability refers to the three questions for perceived quality. Levene's tests looks at equality of variances across the four experimental cells. Shapiro Wilk tests study normality in the residuals from the ANOVA models.

Table 11. Readability check ANOVA

Effect	F	df	p	Partial eta squared
Label style	32.67	1, 121	< 0.001	0.213
Displayed price	0.15	1, 121	0.702	0.001
Label style x displayed price	5.19	1, 121	0.024	0.041

Note. $N = 125$. The dependent variable is the readability value measured on a 1-7 scale. Results are from a 2x2 between-subjects ANOVA using Type III sums of squares. The effect sizes are partial eta squared.

4.3. Hypothesis tests

H1 and H2 were tested with two separate 2x2 between-subjects ANOVAs, with label style and displayed price as independent variables. H3 was tested through the interaction between label style and displayed price in the purchase intention model. The cell means are shown in Table 12, and our ANOVA results are in Tables 13 and 14.

Table 12. Cell means for perceived quality and purchase intention by condition

Condition	n	Perceived quality M (SD)	Purchase intention M (SD)
Traditional at 99 SEK	31	3.65 (1.47)	3.61 (1.78)
Traditional at 199 SEK	30	4.91 (1.38)	3.97 (1.54)
Modern at 99 SEK	32	3.85 (1.69)	4.62 (1.93)
Modern at 199 SEK	32	4.50 (1.77)	4.09 (1.89)

Note. $N = 125$. Values are means with standard deviations in parentheses. Perceived quality is the three questions average; purchase intention is a single item. Both outcomes are measured on 1 to 7 scales.

4.3.1. H1, Effect of label style on perceived quality

We predicted in H1 that the modern label would generate higher perceived quality than the traditional label, and this was not supported. As shown in Table 13, there was no significant main effect of label style on perceived quality, and the small difference between the label styles was not in the predicted direction. Across price conditions, perceived quality was (descriptively) slightly higher for the traditional label than for the modern label, opposite to our prediction. At the cell level, the traditional label at 199 SEK produced the highest perceived quality mean of any condition ($M = 4.91$), above the modern label at 199 SEK ($M = 4.50$). This not significant pattern is consistent with the idea that traditional design cues may align with higher price signals although this pattern cannot be interpreted as a confirmed effect.

Table 13. Two-way ANOVA results for perceived quality

Effect	<i>df</i>	<i>F</i>	<i>p</i>	partial η^2
Label Style	1, 121	0.13	0.723	0.001
Displayed price	1, 121	11.27	0.001	0.085
Label style x Price	1, 121	1.19	0.278	0.010

Note. $N = 125$. Results are from a 2x2 between-subjects ANOVA using Type III sums of squares. The dependent variable is perceived quality. Effect sizes are partial eta squared.

4.3.2. H2, Effect of label style on purchase intention

H2 predicted that the modern label would generate higher purchase intention than the traditional label, and this hypothesis was not supported at the conventional 0.05 significance level. As shown in Table 14, the main effect of label style on purchase intention did not reach statistical significance ($p = 0.079$). The descriptive means were in the predicted direction, with the modern label producing higher purchase intention than the traditional label across price conditions, but this difference should not be treated as statistical support for H2. Displayed price also had no significant main effect on purchase intention.

Table 14. Two-way ANOVA results for purchase intention

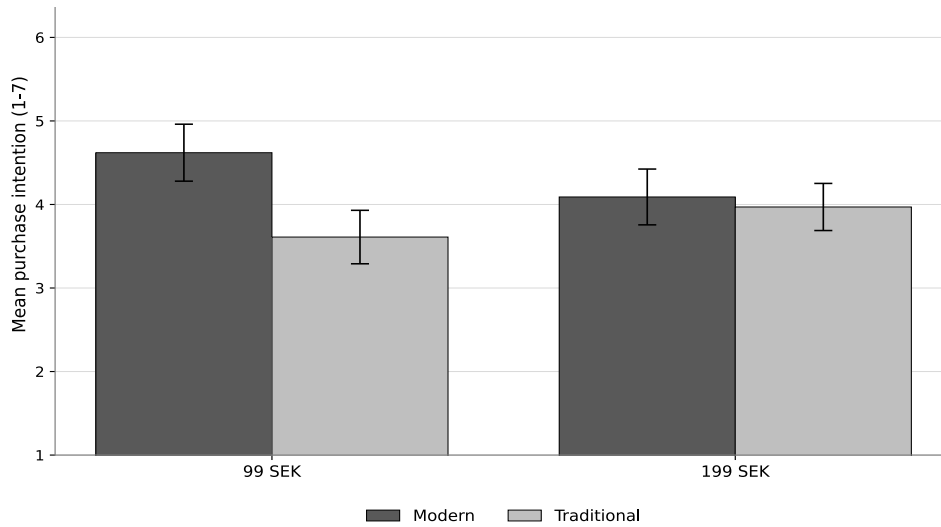
Effect	<i>df</i>	<i>F</i>	<i>p</i>	partial η^2
Label style	1, 121	3.14	0.079	0.025
Displayed price	1, 121	0.08	0.783	0.001
Label style x Price	1, 121	1.89	0.171	0.015

Note. $N = 125$. Results are from a 2x2 between-subjects ANOVA using Type III sums of squares. The dependent variable is purchase intention. Effect sizes are partial eta squared.

4.3.3. H3, Moderating role of displayed price on purchase intention

H3 predicted that displayed price would moderate the effect of label style on purchase intention, so that the purchase intention advantage of the modern label would be smaller at 199 SEK compared to 99 SEK. As shown in Table 14, the interaction between label style and displayed price was not statistically significant, and so H3 was not supported. The cell means did move in our expected direction though, since the modern label showed a purchase intention advantage at 99 SEK, with a smaller difference between label styles at 199 SEK as seen in Figure 6. Because the interaction was once again not statistically significant however, this cannot be treated as evidence that displayed price moderated the label style effect in our study.

Figure 6. Mean purchase intention scores by label style and displayed price (1-7 scale).



Note. Error bars represent a ± 1 standard error of the mean. $n = 31$ (Traditional, 99 SEK), 30 (Traditional, 199 SEK), 32 (Modern, 99 SEK), 32 (Modern, 199 SEK).

4.3.4. Sensitivity analysis controlling for readability

Because the modern label was rated as easier to read than the traditional label, we conducted a main experiment sensitivity analysis with readability as a covariate, testing whether our main conclusions changed after considering a readability difference.

For perceived quality, label style remained not significant, displayed price remained significant, and the interaction between label style and displayed price remained not significant, and the same remained true for purchase intention.

Readability itself showed only weak and non-significant tendencies in both models, and the main conclusions of our experiment did not change after readability was included as a covariate. This should still not be read as a full statistical correction of the readability issue, however, because readability was measured only after respondents had seen the stimulus and is likely shaped by the visual manipulation itself. In other words, readability is not a clean background variable that existed before the manipulation but something that may partly belong to the way the label style was experienced. Including it as a covariate can adjust our model in a useful exploratory sense, but it cannot prove that readability has been fully separated from the label style manipulation. The analysis should therefore be interpreted as a sensitivity check showing that the main results do

not change when readability is added to the model and not as evidence that the readability difference has been fully controlled away. Additional exploratory checks with self-rated wine knowledge, wine purchase frequency, and wine consumption frequency also did not change the interpretation of our main hypothesis tests. These analyses were not central to the conceptual model and were therefore treated as robustness checks rather than additional hypothesis tests in Appendix C.

Table 15. Sensitivity analysis with readability as the covariate

Dependent variable	Effect	F	df	p	partial η^2
Perceived quality					
	Label style	1.17	1, 120	0.281	0.010
	Displayed price	11.03	1, 120	0.001	0.084
	Readability	2.75	1, 120	0.100	0.022
	Label style x displayed price	0.54	1, 120	0.462	0.005
Purchase intention					
	Label style	0.61	1, 120	0.437	0.005
	Displayed price	0.12	1, 120	0.735	0.001
	Readability	3.06	1, 120	0.083	0.025
	Label style x displayed price	1.01	1, 120	0.317	0.008

Note. $N = 125$. Results are from 2x2 models including readability as a covariate. Readability was measured after stimulus exposure on a 1 to 7 scale, so this analysis should be interpreted as an exploratory sensitivity analysis only. P-values are two-sided.

5. Discussion

5.1. Findings

The experiment produced one clear significant model finding and three unsupported hypotheses. Displayed price did significantly increase perceived quality, since our respondents rated the wine as higher quality at 199 SEK than at 99 SEK. Label style, on the other hand, did not significantly affect perceived quality or purchase intention, and displayed price did not significantly moderate the effect of label style on purchase intention. Our H1, H2, and H3 were not statistically supported.

Displayed price worked as a quality cue, while the label style comparison did not reach statistical significance on either outcome. The observed label style effect on purchase intention returned a value for p of 0.079, just outside the conventional 0.05 threshold, and the descriptive direction matched the prediction. With approximately 31 respondents per cell, the design had limited power to detect small effects, and a result close to the threshold may therefore reflect either a small underlying effect or sampling variation, and the present data cannot distinguish between these interpretations. When readability was included as a covariate in the sensitivity analysis (Section 4.3.4), the label style effect on purchase intention became slightly weaker, suggesting that the observed pattern may partly reflect the readability difference between stimuli rather than label style alone. The H2 pattern should therefore be treated as a candidate for replication of higher power with stimuli matching in readability and not as evidence of a confirmed effect of label style. H2 and H3 are also linked by design, since H3 predicted that the modern label's purchase-intention advantage would be smaller at the higher price. The non-significant interaction leaves this possibility open but once again does not provide evidence that displayed price moderated the label style effect.

5.2. Relation to prior literature.

The price to quality effect is consistent with the cue literature we explored. Dodds et al. (1991) reported that price affected perceived quality and willingness to buy in different ways, and our results show a similar split where price affects perceived quality but not purchase intention. Völckner and Hofmann (2007) note that the strength of the price to

quality link varies by product category, study design, and country, and our result adds a new case (wine evaluated in a regulated Swedish monopoly setting) in which the relationship holds.

The label style results connect to a more divided literature, since on one side Lacoste-Badie and Droulers (2024) found that modern alcohol labels attracted more attention and were chosen more often than traditional alternatives in a comparative shelf-choice setting, but on the other, Boudreaux and Palmer (2007) found traditional labels linked to higher perceived value and an upper-class brand personality, with Celhay and Passebois (2011) especially finding that typicality preference strengthens under perceived risk. Our not significant result for H1 aligns more closely with the second type of research, and the means we calculated run in that direction. Our result is likely not an anomaly but perhaps a case the H1 prediction did not anticipate.

The MAYA effect documented by Hekkert et al. (2003) was tested in comparison contexts, where a stimulus is judged against a typicality baseline. Our focal setting gives no baseline of this type, so the lack of statistical support is consistent with the possibility that "MAYA" novelty effects need some form of comparison context, which would also help explain why Lacoste-Badie and Droulers (2024) found a modern label advantage in a choice setting that the present focal bottle design does not have.

The H3 interaction was not statistically supported, so there is no confirmed result to relate to prior literature. The pattern we saw, with the smallest modern-over-traditional gap at the higher price, is consistent with the price-tier reading of traditional cues linking traditional design codes to upper-class impressions but our H3 results leave essentially the question open instead of contradicting established findings.

5.3. Theoretical contribution

Our contribution is smaller than our conceptual model originally envisioned but is specific. The clearest contribution is empirical, since the price to quality cue relationship held inside a regulated Swedish retail context where promotional signals are constrained and displayed price is one of few visible cues available to the consumer. The study also separates perceived quality from purchase intention and shows that the two outcomes did not respond in the same way, i.e. displayed price affected perceived quality but not purchase intention, while label style showed only a weak not significant tendency for purchase intention and no effect on perceived quality. This suggests that future similar wine label work should avoid reducing evaluation and intention down to a single general response measure.

A further framing choice is our focal bottle approach, since rather than studying label only or a full comparative shelf, our experiment isolates a stage in which one bottle has already become object of evaluation. This narrows the scope of the claims but also clarifies what the study can and cannot explain, since we speak to the evaluation of a displayed bottle and not to attention, consideration set formation, or actual purchase behaviour for instance. The proposed cue fit moderation between displayed price and label style was not supported. This should once again be treated as a theoretical direction for future research instead of direct contribution.

5.4. Practical implications

The practical implications should be interpreted cautiously because the label-style hypotheses were not statistically supported. The clearest practical implication concerns displayed price, since in this study the 199 SEK price increased perceived quality relative to the 99 SEK price, suggesting that displayed price can function as a quality signal in a Systembolaget-like focal-bottle evaluation. For label design, the implication is not that label design is unimportant in general. This study does not provide evidence that changing between the selected modern and traditional label styles is a reliable way to increase perceived quality or stated purchase likelihood once a bottle is already the focus of evaluation. If the managerial objective is specifically to raise perceived quality or purchase likelihood for a bottle already considered, then label style should be treated

as a lower-priority lever than price positioning based on this study's evidence. This conclusion should not be extended too far as label design may still matter for attention, shelf standout, brand positioning, and entry into the consumer's consideration set, but those stages were not tested here. The practical recommendation is therefore conservative and can be worded as a consideration of price and label cues together as well as testing of label effects in a fuller shelf-choice design before making strong managerial design recommendations.

5.5. Limitations

Our study has several limitations. Purchase intention was measured with a single item, which is defensible for concrete singular constructs (Bergkvist & Rossiter, 2007) but does not allow for any internal consistency estimation. The scenario also asked respondents to imagine the bottle had already become a candidate under consideration. This was of course deliberate and let us focus on the late evaluation stage, but the experiment cannot speak to earlier stages such as how labels attract attention on a busy shelf or how a wine enters the consideration set in the first place. The fully between-subjects design was chosen to avoid demand effects from a within-subjects presentation of both label styles, but our choice reduced statistical power. With around 31 respondents per cell our analysis had limited sensitivity to detect small main effects and small/moderate interactions. The pattern close to significance for H2 ($p = 0.079$) and the directionally consistent (but clearly not significant) interaction in H3 are both consistent with an underpowered design, not necessarily a study without effects. A larger replication would be needed to choose between these interpretations.

Stated intentions do not always translate into actions, since circumstances such as access, available money, or timing can change between the survey and the actual shelf encounter (Ajzen, 1991). This is particularly relevant for wine, where purchases are often tied to specific occasions and the final decision can shift away from a stated intention. The sample was recruited through our social media and personal networks and was therefore heavily skewed toward male respondents aged 20 to 23. Respondents were also predominantly familiar with Systembolaget, which supports the internal interpretability of the shelf scenario but means the findings should not be generalised to

consumers without Systembolaget experience (who may rely on our provided cues differently in other countries).

Only two price points were tested, both within recurring Systembolaget price ranges reported in Appendix B. This means that the study cannot show how label style would interact with clearly budget or clearly premium prices, and the higher price was also rated as somewhat less fitting for the wine/Systembolaget context than the lower price, although the mean remained above the scale midpoint. Our scenario also framed the price as part of the respondent's stated budget, which may have made both price levels appear more acceptable and may at least in part explain why displayed price affected perceived quality but not purchase intention. In other words, our own design likely reduced the role of price as a cost cue.

Our study also covers exclusively wine, not beer or other beverages/spirits. Visual cues may be interpreted differently across categories because category conventions, expected price ranges, and the meanings of design styles are not the same in beer or spirits as they are in wine. Findings from this study should therefore not be extended beyond wine without further testing in the relevant categories. The stimuli were also fictional, which helped us avoid brand familiarity effects, but it also means that we did not capture the role of real producer reputation in actual wine decisions. The study examines label and price cues inside a simulated regulated Swedish retail context shaped by the Systembolaget monopoly and its public health mission without promotion (Alkohollag, 2010:1622). Because it does not compare Systembolaget with another retail setting, the context should be treated as a scope condition rather than a causal explanation. Our findings should therefore be described as specific evidence from a simulated focal bottle evaluation context only.

Finally, the modern label was rated as more readable than the traditional label. This readability gap is an important confound for the label style results, and while a sensitivity analysis including readability as a covariate did not change the pattern of significance for any hypothesis, this did not resolve the confound. Readability was measured after stimulus exposure and is plausibly downstream of the visual manipulation rather than independent from it. Including a variable following the

treatment as a covariate can bias effect estimates instead of correcting them, so the adjusted analysis is descriptive only. The honest conclusion is that readability remains a confound this study cannot fully account for, and future work should match pilot stimuli on readability before proceeding with a main experiment.

5.6. Future research

The most direct extension of this thesis would be a larger sample replication of the same 2x2 design. A larger sample would make it possible to estimate the label style and interaction effects with more precision, and it would help explain whether the directional purchase intention pattern observed here reflects a small but real effect or only sampling variation. This type of replication should keep the same basic differentiation between label style and displayed price, but it should ideally include a more balanced sample across age groups and wine knowledge levels. Future research could also examine the earlier stage of consideration and choice set formation (Shocker et al., 1991). A shelf choice study with several competing bottles could test whether modern or traditional labels are more likely to enter the consumer's consideration set before final evaluation, complementing our study which focuses on evaluation of one focal candidate rather than on the selection from a full shelf.

A further extension would be to use behavioural outcomes. In an online experiment, purchase intention is the most realistic dependent variable, but future studies could measure click choice, bottle selection, simulated “add to basket”, actual pickup behaviour, or purchase behaviour where this is practically (and ethically) possible. This would make it easier to evaluate if the patterns found in stated intention also appear when consumers make more concrete decisions. Future research should also test a wider price range because our cue fit argument would be better tested by including lower budget prices, middle everyday prices, and clearly premium prices. If traditional design bundles become more persuasive when the price signal becomes more premium, this should become clearer when the price range is larger than in our experiment.

Finally, our Systembolaget context should be compared with less regulated retail environments and countries, since a similar experiment in a conventional supermarket/wine store, or online store could show whether the same pattern appears

when promotional materials, brand competition, discounts, or retailer cues are more visible. This would help separate what is specific to the Swedish regulated retail setting from what can be generalised to wine retail more broadly abroad.

6. Conclusion

Our thesis examined whether modern versus traditional wine label design and displayed price affect perceived quality and purchase intention when consumers evaluate one focal bottle in a Systembolaget-like shelf context. In a 2x2 between-subjects experiment with 125 respondents, displayed price significantly increased perceived quality, while label style produced no significant effects on both outcomes and displayed price did not significantly moderate the label style effect on purchase intention. The H2 main effect on purchase intention was consistent with our prediction direction but also not statistically significant ($p = 0.079$).

Our defensible contribution is narrower than the conceptual model originally implied. We replicate the price-to-perceived-quality cue effect inside of a regulated Swedish retail frame that prior wine label work has not used in this way, and also document that perceived quality and purchase intention do not respond to the same cues in our data, which argues against using the two outcomes into one general response measure in future label research. Our focal bottle framing is a scope condition that defines what we can and cannot speak to instead of a real contribution in and of itself.

The use of one selected label per style and an unresolved readability difference between our stimuli are of course caveats that impact what our evidence can carry. A replication of higher power with a more balanced sample in terms of age and stimuli matching in readability are the most direct next steps. What our study establishes is smaller scale but specific in that in a simulated Systembolaget-like focal bottle evaluation, displayed price functioned as a quality cue for an otherwise identical wine, while label style did not.

Appendix

Appendix A. Consideration set decision framework

Figure 7. Consideration set theory illustration. *Reproduced from Shocker et al. (1991, p. 184).*

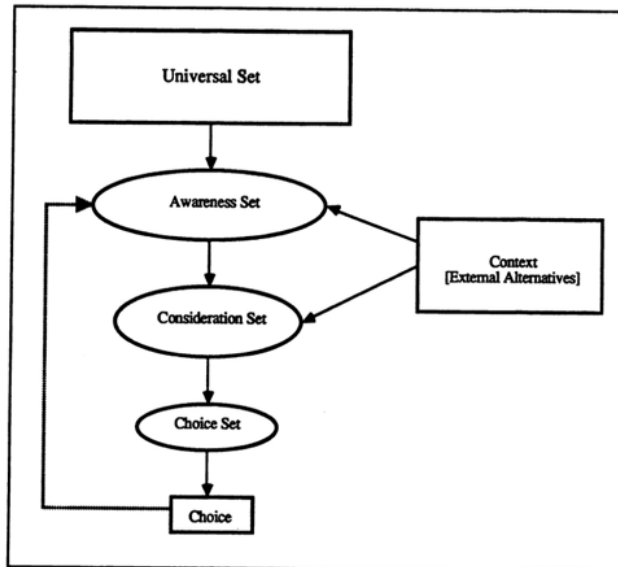
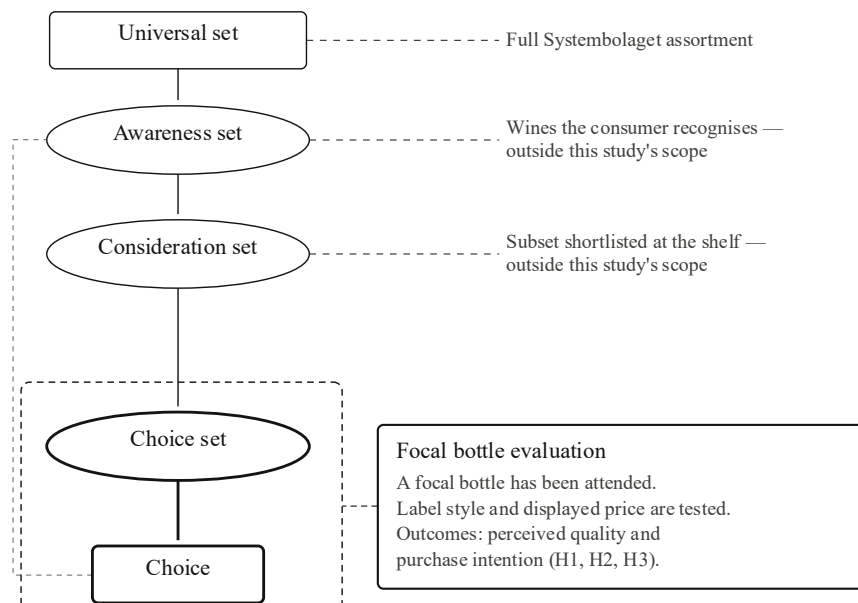


Figure 8. The consideration set framework applied to the Systembolaget shelf context.



*Adapted from Shocker (1991).
Context (external alternatives) omitted from the original for clarity.*

Appendix B. Justification of experimental price levels

Figure 9. Sample from Systembolaget 09/2026 launch plans

2026/06	RÖTT VIN	2302a RÖTT VIN BOX > 2L (8
2026/06	RÖTT VIN	
2026/06	RÖTT VIN	2104a RÖTT VIN (80-99)
2026/06	RÖTT VIN	2108 RÖTT VIN (150-199)
2026/06	VITT VIN	3106 VITT VIN (100-119)
2026/06	VITT VIN	3104a VITT VIN (80-99)
2026/06	VITT VIN	3104a VITT VIN (80-99)
2026/06	VITT VIN	3106 VITT VIN (100-119)
2026/06	VITT VIN	3302a VITT VIN BOX > 2L (76
2026/06	VITT VIN	3110 VITT VIN STORA FLASK

The two prices used in our experiment, 99 SEK and 199 SEK, were not selected at random, but were taken from Systembolaget's public launch plans available on their website (the discretionary choice used here as example is the sample month of September 2026). The plans divide the red and white high volume wine categories into price ranges that reappear at every launch. The figure above shows the rows from the launch plan we drew on, with the circled areas marking the top of lowest range (99 SEK) and the top of higher range used in our study (199 SEK). We picked the top of each range so that each price would line up with a boundary/customer segmentation Systembolaget itself uses, while still leaving a clear gap between the two conditions. Both prices fall inside the same generic wine category, which means the price manipulation moves within Systembolaget's own found customer structure instead of across it.

Appendix C. Robustness checks: ANCOVA with behavioral covariates

We conducted some ANCOVA robustness checks in R by adding wine knowledge, wine purchase frequency and wine consumption frequency as covariates to the label style x price models from the thesis. These exploratory covariates did NOT alter the pattern of the main hypothesis tests. For perceived quality, the price effect remained significant, and label style and the interaction remained not significant. Likewise, for purchase intention, label style, price, and their interaction remained not significant. None of the three exploratory covariates reached statistical significance in any robustness model and the R Studio output is here attached.

Figure 10. Robustness ANCOVA output: perceived quality.

```
=====
Robustness ANCOVA for outcome: perceived_quality
=====
Anova Table (Type III tests)

Response: perceived_quality
              Sum Sq  Df  F value  Pr(>F)
(Intercept)    2234.71  1 895.6181 < 2e-16 ***
label_style      2.36   1   0.9456 0.33283
price          27.05   1  10.8417 0.00131 **
wine_knowledge_c  0.04   1   0.0157 0.90037
purchase_frequency_c 4.91   1   1.9672 0.16337
consumption_frequency_c 0.13   1   0.0525 0.81915
label_style:price  3.27   1   1.3093 0.25484
Residuals      294.43 118
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Partial eta-squared:
# Effect Size for ANOVA (Type I)

Parameter          | Eta2 (partial) | 95% CI
-----
label_style         | 8.71e-04 | [0.00, 1.00]
price               | 0.09 | [0.02, 1.00]
wine_knowledge_c    | 0.02 | [0.00, 1.00]
purchase_frequency_c | 0.02 | [0.00, 1.00]
consumption_frequency_c | 6.22e-05 | [0.00, 1.00]
label_style:price   | 0.01 | [0.00, 1.00]

- One-sided CIs: upper bound fixed at [1.00].
Model R-squared:
[1] 0.1283031
Adjusted R-squared:
[1] 0.0839795
```

Figure 11. Robustness ANCOVA output: purchase intention.

```
=====
Robustness ANCOVA for outcome: purchase_intention
=====
Anova Table (Type III tests)

Response: purchase_intention
              Sum Sq  Df  F value Pr(>F)
(Intercept)    2076.53  1 666.8159 <2e-16 ***
label_style      2.97  1  0.9522 0.3312
price           0.46  1  0.1490 0.7002
wine_knowledge_c 1.33  1  0.4267 0.5149
purchase_frequency_c 5.91  1  1.8973 0.1710
consumption_frequency_c 0.03  1  0.0090 0.9248
label_style:price 5.55  1  1.7808 0.1846
Residuals      367.46 118
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Partial eta-squared:
# Effect Size for ANOVA (Type I)

Parameter          | Eta2 (partial) | 95% CI
-----
label_style         | 0.03 | [0.00, 1.00]
price               | 8.40e-04 | [0.00, 1.00]
wine_knowledge_c    | 0.04 | [0.00, 1.00]
purchase_frequency_c | 0.03 | [0.00, 1.00]
consumption_frequency_c | 4.65e-05 | [0.00, 1.00]
label_style:price   | 0.01 | [0.00, 1.00]

- One-sided CIs: upper bound fixed at [1.00].
Model R-squared:
[1] 0.09758652
Adjusted R-squared:
[1] 0.05170109
```

Appendix D. AI tools declaration

We now address the four disclosure questions required by the Stockholm School of Economics guidelines found on Canvas. The addressed four required questions are (1) What AI tools have been used and how? (2) In what ways have these tools contributed to increasing the quality of the thesis? (3) What potential risks were found using AI and what measures were taken to reduce these risks? (4) What insights were gained from using AI tools in the thesis writing process?

Two tools were used, namely ChatGPT 5.3 (OpenAI), accessed through the HHS dedicated workspace, and Claude Sonnet 4.6 (Anthropic) on the Teams plan. Each author worked with one of these tools throughout the project.

Both tools were used under a custom markdown prompt that we wrote assisted by AI, specifying the limitations and use of AI in the thesis. The prompt mostly limited the chatbots to critical reading instead of writing. It was prohibited from producing final text, introducing academic sources we had not already read and linked to our claims, or modifying all the claims we had verified manually in our text. The prompt was updated once near the end of the process to include spelling checks on pasted paragraphs before final submission. Our main use was pasting our own handwritten text into the chat window for feedback on word choice and grammar/sentences over time as well as internal logic and argumentation. Sample prompts from our work included similar questions:

- "Here is our draft of section 2.2 on visual semiotics, we are not sure if the favier material on design complexity belongs here as we have built or if its better as a separate construct since it confuses in part what the thesis manipulates. Is this section consistent? flag any internal logical problems you identify"
- "Check my final thesis draft in the project files named 'Working Draft FINAL' with my own comments in it as well. The goal now is to make it easier to read with additional tables instead of excessive numbers in text where it could help the reader by improving the flow, so analyse it and identify where our numbers in text benefit from a proper table"

The chatbots also helped us understand documentation for R functions in our ANOVA implementation. We ran the analysis ourselves on our local machines, and the chatbots served helped interpreting what specific commands did in R Studio. The ANCOVA robustness code in Appendix C was AI generated on our existing and working R project. We also used AI to suggest candidate items for the instruction validity check in our survey (the options for wine labels, beer cans, pizza packaging, and coffee packaging). Reviews of our full thesis draft were conducted late in the process, focused mainly on identifying where exactly in our text we could benefit from illustrations. Beyond this, we compiled personal handwritten lecture notes from the quantitative track into an AI summary for our own structure early on, but that material obviously never appeared in the thesis directly. All academic sources were found and read by us through ProQuest and HHS library resources, managed via a shared Zotero folder. On one occasion we pasted our reference list into the chatbot to verify that two specific Systembolaget sources and the Alcohol Act citation were formatted and integrated correctly in APA. The structure of this disclosure was also shaped with AI help, since past marketing theses at the department used a similar format and we wanted to avoid our version looking too similar to other students' work. No other AI tools, grammar checkers, or translation software were used.

2. How did these tools contribute to the quality of the thesis?

The chatbot setup forced us to defend our own reasoning to a very large extent. Since the tools were asked to challenge weak logic and unsupported claims all the time, we had to justify choices we were taking in the text to ourselves more clearly than we might have done otherwise in real time. This also helped produce more consistent text across sections written by two different people in terms of language.

3. What risks were identified, and how were they handled?

Despite clear instructions not to do so, both tools occasionally entirely made up references or shifted the meaning of claims we had provided, sometimes even after being corrected. This really made us appreciate the restriction rules in our system prompt and reminded us of treating every output as input to our own thinking instead of likely incorrect ready conclusions. All sources were located and verified independently

by us through the school's academic databases. We did not upload full article PDFs or course material into either tool, both under HHS guidelines and out of copyright issues. Survey data exported from Qualtrics contained no personally identifiable information, which we confirmed before uploading and using on our machines.

4. What did we learn from using AI in this process?

The AI worked best when we approached bots with concrete decisions to work through, especially with regards to style and word choice. Despite continued improvements across generations of AI models, chatbot outputs remain unreliable enough in logic and facts that our verification by hand was always necessary at every step of the writing process despite the AI help with e.g. word choices or past/present verbs' consistency.

References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Alkohollag [Alcohol Act] (SFS 2010:1622).
Riksdag. https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/alkohollag-20101622_sfs-2010-1622/
- Barber, N., Almanza, B. A., & Donovan, J. R. (2006). Motivational factors of gender, income and age on selecting a bottle of wine. *International Journal of Wine Marketing*, 18(3), 218–232. <https://doi.org/10.1108/09547540610704774>
- Bergkvist, L., & Rossiter, J. R. (2007). The Predictive Validity of Multiple-Item versus Single-Item Measures of the Same Constructs. *Journal of Marketing Research*, 44(2), 175–184. <https://doi.org/10.1509/jmkr.44.2.175>
- Bloch, P. H. (1995). Seeking the Ideal Form: Product Design and Consumer Response. *Journal of Marketing*, 59(3), 16–29. <https://doi.org/10.1177/002224299505900302>
- Boudreaux, C. A., & Palmer, S. E. (2007). A charming little Cabernet: Effects of wine label design on purchase intent and brand personality. *International Journal of Wine Business Research*, 19(3), 170–186. <https://doi.org/10.1108/17511060710817212>
- Celhay, F., Masson, J., Garcia, K., Folcher, P., & Cohen, J. (2017). Package graphic design and innovation: A comparative study of Bordeaux and Barossa wine visual codes. *Recherche et Applications En Marketing (English Edition)*, 32(2), 46–70. <https://doi.org/10.1177/2051570716685524>
- Celhay, F., & Passebois, J. (2011). Wine labelling: Is it time to break with tradition? A study of the moderating role of perceived risk. *International Journal of Wine Business Research*, 23(4), 318–337. <https://doi.org/10.1108/17511061111186497>
- Celhay, F., & Remaud, H. (2018). What does your wine label mean to consumers? A semiotic investigation of Bordeaux wine visual codes. *Food Quality and Preference*, 65, 129–145. <https://doi.org/10.1016/j.foodqual.2017.10.020>
- Celhay, F., & Trinqucoste, J. F. (2015). Package Graphic Design: Investigating the Variables that Moderate Consumer Response to Atypical Designs. *Journal of Product Innovation Management*, 32(6), 1014–1032. <https://doi.org/10.1111/jpim.12212>
- Perdue, B. C., & Summers, J. O. (1986). Checking the Success of Manipulations in Marketing Experiments. *Journal of Marketing Research*, 23(4), 317–326. <https://doi.org/10.1177/002224378602300401>
- Dodd, T. H., Laverie, D. A., Wilcox, J. F., & Duhan, D. F. (2005). Differential Effects of Experience, Subjective Knowledge, and Objective Knowledge on Sources of Information used in Consumer Wine Purchasing. *Journal of Hospitality & Tourism Research*, 29(1), 3–19. <https://doi.org/10.1177/1096348004267518>
- Dodds, W. B., Monroe, K. B., & Grewal, D. (1991). Effects of Price, Brand, and Store Information on Buyers' Product Evaluations. *Journal of Marketing Research*, 28(3), 307–319. <https://doi.org/10.1177/002224379102800305>
- Greco, L. M., O'Boyle, E. H., Cockburn, B. S., & Yuan, Z. (2018). Meta-Analysis of Coefficient Alpha: A Reliability Generalization Study. *Journal of Management Studies*, 55(4), 583–618. <https://doi.org/10.1111/joms.12328>
- Hekkert, P., Snelders, D., & Van Wieringen, P. C. W. (2003). 'Most advanced, yet acceptable': Typicality and novelty as joint predictors of aesthetic preference in

- industrial design. *British Journal of Psychology*, 94(1), 111–124.
<https://doi.org/10.1348/000712603762842147>
- Lacoste-Badie, S., & Droulers, O. (2024). How alcohol label design influences attention among young adults: Evidence from an eye-tracking study. *Alcohol and Alcoholism*, 60(1), agae090. <https://doi.org/10.1093/alcalc/ague090>
- Lockshin, L., & Corsi, A. M. (2012). Consumer behaviour for wine 2.0: A review since 2003 and future directions. *Wine Economics and Policy*, 1(1), 2–23.
<https://doi.org/10.1016/j.wep.2012.11.003>
- Lockshin, L. S., & Timothy Rhodus, W. (1993). The Effect of Price and Oak Flavor on Perceived Wine Quality. *International Journal of Wine Marketing*, 5(2/3), 13–25.
<https://doi.org/10.1108/eb008614>
- Orth, U. R., & Malkewitz, K. (2008). Holistic Package Design and Consumer Brand Impressions. *Journal of Marketing*, 72(3), 64–81.
<https://doi.org/10.1509/JMKG.72.3.064>
- Öztürk, B., & Ertamay, S. İ. (2019). Wine bottle design thinking modelling: An analysis of local wine brands within Urla vineyard road. *BIO Web of Conferences*, 15, 03014.
<https://doi.org/10.1051/bioconf/20191503014>
- Sheeran, P. (2002). Intention—behavior relations: A conceptual and empirical review. *European Review of Social Psychology*, 12(1), 1–36. <https://doi.org/10.1080/14792772143000003>
- Sherman, S., & Tuten, T. (2011). Message on a bottle: The wine label's influence. *International Journal of Wine Business Research*, 23(3), 221–234.
<https://doi.org/10.1108/17511061111163050>
- Shocker, A. D., Ben-Akiva, M., Boccara, B., & Nedungadi, P. (1991). Consideration Set Influences on Consumer Decision-Making and Choice: Issues, Models, and Suggestions. *Marketing Letters*, 2(3), 181–197. <https://doi.org/10.1007/BF00554125>
- Söderlund, M. (2018). *Experiments in marketing* (First edition). Studentlitteratur.
- Spears, N., & Singh, S. N. (2004). Measuring Attitude toward the Brand and Purchase Intentions. *Journal of Current Issues & Research in Advertising*, 26(2), 53–66.
<https://doi.org/10.1080/10641734.2004.10505164>
- Spence, C. (2024). Cognitive influence on the evaluation of wine: The impact and assessment of price. *Food Research International*, 187, 114411.
<https://doi.org/10.1016/j.foodres.2024.114411>
- Steenkamp, J.-B. E. M. (1990). Conceptual model of the quality perception process. *Journal of Business Research*, 21(4), 309–333. [https://doi.org/10.1016/0148-2963\(90\)90019-A](https://doi.org/10.1016/0148-2963(90)90019-A)
- Systembolaget (2026). *Vårt uppdrag*. Available at:
<https://www.omsystembolaget.se/vart-uppdrag/>
- Systembolaget (2026). *Lanseringsplan september 2026* [Launch plan, September 2026, Excel file]. Retrieved March 24, 2026. Available at: <https://cms-cdn.systembolaget.se/4a7fla/globalassets/nytt/om-vara-nyheter/lanseringsplaner/excel-och-pdf/lanseringsplan-september-2026.xlsx>
- Völckner, F., & Hofmann, J. (2007). The price-perceived quality relationship: A meta-analytic review and assessment of its determinants. *Marketing Letters*, 18(3), 181–196. <https://doi.org/10.1007/s11002-007-9013-2>
- Warshaw, P. R., & Davis, F. D. (1985). Disentangling behavioral intention and behavioral expectation. *Journal of Experimental Social Psychology*, 21(3), 213–228. [https://doi.org/10.1016/0022-1031\(85\)90017-4](https://doi.org/10.1016/0022-1031(85)90017-4)

Zeithaml, V. A. (1988). Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence. *Journal of Marketing*, 52(3), 2–22.
<https://doi.org/10.1177/002224298805200302>