ANIMAL SPIRITS:
EXAMINING STOCK MARKET IRRATIONALITY VIA EFFECTS OF LANGUAGE

BY

XI LIU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois Urbana-Champaign, 2022

Urbana, Illinois

Doctoral Committee:

Professor Dov Cohen, Chair
Professor Dolores Albarracín
Assistant Professor Joey Cheng
Professor Jeff Loewenstein
Professor Li-Jun Ji

ABSTRACT

Four studies investigated grammatical language effects in a high-stakes situation: stock investment. The grammatical gender of the company name in Spanish (Studies 1 - 2) and the classifier used to describe the company in Chinese (Studies 3 - 4) were shown to affect individuals' prediction of future stock movement both in the real world (Studies 1 & 3) and in lab experiments (Studies 2 & 4). The language effect depended in part on the direction of the stock's past movement. In the Spanish grammatical gender studies, upward momentum was consistently greater for companies with grammatically male names than those with grammatically female names. Experimental manipulation suggested that this effect can be plausibly driven by the belief in males' greater forcefulness and rationality, as shown by the negative comments about a lack of the two qualities casting a greater influence on the female stocks. There was also mixed evidence for a surprising finding about male stocks being more overvalued than female stocks. In the Chinese classifier studies, effects were seen after downward trends: companies described with an animal classifier Zhi were more likely to be expected to reverse a downward trend and go back up compared to companies described with an object classifier Zhi. The different meanings of agency in male vs. female and in animal vs. inanimate beings were discussed.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1
CHAPTER 2: SPANISH GRAMMATICAL GENDER STUDIES 15
CHAPTER 3: CHINESE CLASSIFIER STUDIES 46
CHAPTER 4: GENERAL DISCUSSION 76
REFERENCES 80
APPENDIX A. EXAMPLES OF COMPANY NAMES IN STUDY 1 88
APPENDIX B. STUDY 2 NEWS, COMMENTS, AND GRAPHS 89
APPENDIX C. PILOT STUDY 3.2 92
APPENDIX D. STUDY 3 EXAMPLES OF VERBS 93
APPENDIX E. STUDY 4 PRIMING SENTENCES 94
APPENDIX F. STUDY 1 DESCRIPTIVE STATS FOR THE NEUTRAL CATEGORY 95

CHAPTER 1: INTRODUCTION

1.1 Language and Thought

The relationship between language and thought has attracted great research interest over the years. While few would disagree that language is a product of thought, much more debate revolves around whether, and to what extent, thought is influenced by language. Today, a deterministic view of the thought-language relationship has become rather obsolete (Heider, 1972; Li & Gleitman, 2002; Lindsey et al, 2015). However, much cross-cultural evidence supports the weak form of the Sapir-Whorf hypothesis, arguing that language does at least have some influence over the way that we perceive, understand, and remember (Hunt & Agnoli, 1991; Sera, Berge, & del Castillo, 1994; Slobin, 1996; Roberson, Davies, & Davidoff, 2002).

For example, the same concept that could be easily expressed with a single word in one language may require much elaboration in another language. This difference may be a reflection of the divergence in historical thinking habits of different language speakers but may also, in turn, *shape* each speakers' habits of thinking (Hunt & Agnoli, 1991), such that those images, feelings, or ideas that are hard to put into words get thought about less often. In addition to ease and availability, *how* something is presented in a language - the words that are typically paired with it when it is mentioned and the context that it is usually mentioned in - could influence how it is mentally construed as well (Slobin, 1996). Take the concept '*time'* as an example. In Chinese, time is often paired with prepositions "above" and "below" rather than "last" and "next" ("last week" would be "the week above", for example). Consequently, Chinese participants tend to visualize time as a vertical medium as opposed to a horizontal one (Boroditsky, 1999), showing a stark contrast with that of English speakers. Lastly, speakers of different languages may also perceive the *relationship* between concepts differently. For

example, a single word (e.g.: stack) could have multiple meanings (e.g.: *chimney*, or *pile*) in language A but not in language B, so speakers of language A may find it easier to draw connections between, say, chimney and pile, than speakers of language B. Similar logic applies to linguistic features that are unique to cultures. For instance, Spanish speakers may find similarities between two words because they share the same grammatical gender; Chinese speakers may group two words together because they share the same classifier. Grammatical gender and classifier are two key linguistic features that will be explored in-depth in the current project.

1.2 Grammatical Gender

Languages fall into three basic groups based on how much of a role *gender* plays (Prewitt-Freilino, Caswell, & Laakso, 2012). *Grammatical gender languages* rank as the most gendered languages, where all the nouns are assigned a gender. Some grammatical gender languages have two genders: masculine or feminine (such as French). Some have a third neuter gender (such as German). The gender of the noun is reflected repeatedly in the grammatical structure of a sentence, such as in the choice of articles (e.g. “le” and “la” in Spanish) and the conjugation of adjectives (e.g. in Spanish, most adjectives end in “o” when describing a masculine noun and end in “a” when describing a feminine noun). Ranking second on the list are the *natural gender languages*, where the natural gender of the object is shown through pronouns (e.g. “he” and “she” in English), but not through any other part of the sentence. Also, inanimate objects or concepts are not gendered. Lastly, some languages do not consistently distinguish gender at all, not even in pronoun choices. Thus, it is grammatically correct to refer to men, women, dogs, and books alike with the same pronoun.

Grammatical gender ought to be distinguished from biological gender, since they do not always converge, not to mention that grammatical gender is also assigned to inanimate objects which, by definition, do not have a biological gender. A quote from Mark Twain neatly illustrates how arbitrary and meaningless gender assignments can be sometimes: "In German, a young lady has no sex, while a turnip has… [A] tree is male, its buds are female, its leaves are neuter. Horses are sexless, dogs are male, cats are female… tomcats included" (2007). Moreover, across languages (especially across languages with different linguistic origins), the gender of inanimate objects can be surprisingly different, leaving the poor language learners bewildered. For example, the word "sun" is masculine in Spanish, feminine in German, and neuter in Russian (Boroditsky, Schmidt, and Phillips, 2003). This phenomenon again shows how arbitrary the grammatical gender categories can be sometimes.

Although arbitrary, grammatical gender may be no less meaningful in the eyes of the speakers: objects with a certain grammatical gender are considered – overtly and subconsciously – to be carrying some corresponding gender-specific qualities. For example, grammatically female objects are more easily remembered to have a female name (Boroditsky et al, 2003), are considered more similar to each other than to grammatically male objects (Phillips & Boroditsky, 2003), and are considered as having more stereotypically feminine traits (Konishi, 1993). The same goes for grammatically male objects. When biological gender and grammatical gender contradicted, as in the case of a male giraffe in German where all giraffes are grammatically feminine, the discrepancy notably interfered with German speakers' cognitive processing as reflected in their longer response times and higher error rates (Imai, Schalk, Saalbach, & Okada, 2014). These findings indicate that although the assignment of grammatical gender does not always follow intuition (Saalbach, Imai, & Schalk, 2012; Boroditsky & Schmidt, 2000), it does

become “meaningful” in a way to the speakers. However, there has also been evidence on the opposite side, suggesting that the grammatical gender effect is not robust to replication (Mickan et al., 2014) or not found in the real world using a Big Data approach to corpora (Kann, 2019). Conceptually also, Sato et al. (2016) make an important distinction between Whorfian effects in which language affects nonlinguistic cognition vs. a thinking-for-speaking approach (Slobin, 1996) that implies effects might occur only during “language-mediated cognition”. The present studies cannot address this important distinction because they examine real-world behavior that almost surely is language-mediated. The present studies can however potentially address another major criticism of the existing evidence about the grammatical gender effect, namely that the effect is an artifact of demand characteristics where participants guessed the experimenters’ intention about classifiers and played along in order to be ‘good participants’ (Sato et al, 2016). The present studies are particularly useful because they examine the linguistic effect in a situation where people have a strong incentive to be correct and will likely not be driven by laboratory demand characteristics.

1.3 Classifier

Another arbitrary grammatical feature that may exert influence on the semantic understanding of a concept and its perceived relationship with others is the classifier. It is a unique grammatical feature that is present in many Asian languages such as Chinese, Japanese, Thai, and Vietnamese (Norman, 1988; Zhang, Shu, & Liu, 2011). They are also commonly referred to as *numeral classifiers, measure words,* or *measures*. These names are descriptive of the classifier’s function in a sentence: typically indicative of the *type*, or *class*, of the noun (Norman, 1988). Classifiers are obligatory in a noun phrase or a sentence that contains numerals

or determinatives. For example, in Mandarin Chinese[1] a determiner or a quantifier cannot directly precede a noun, and a classifier (CL) must be inserted in between to make the sentence grammatically correct (e.g. this CL chicken, five CL chickens) (Tzeng, Chen, & Hung, 1991). In English, the closest analogy to a classifier would be a measure word. For example, the word *glass* in the phrase "a glass of water" takes on a similar role as the classifier in "a CL chicken". However, measure words in English are only obligatory for a small group of nouns (e.g. uncountable nouns) while classifiers in Chinese are required regardless of the noun type.

There are many different classifiers in Chinese. Classifiers are usually used to reflect some salient perceptual properties of the associated noun such as its material, shape, consistency, or size (Allan, 1977). Nouns that share similar perceptual qualities often share the same classifier. The pairing between classifiers and nouns is not one-to-one, but many-to-many. Many nouns could share a classifier. For example, table, photo, and paper all go with the classifier *zhang* because they all usually have a flat surface. Meanwhile, multiple classifiers could be considered grammatically correct choices for the same noun based on the context, writing habit, or the aspect that the speaker/writer would like to emphasize. For example, *zhi*, *tou,* and *shan* could all modify the word *pig*, but *zhi* indicates it is a skinny pig or a piglet, *tou* indicates it is a bulky, muscular animal, while *shan* refers to a dead pig and emphasizes its use as a meat source. Another example: "Yi [one] tiao [classifier] yu [fish]" and "Yi wei [classifier] yu" both mean one fish. However, the classifier *tiao* indicates the elongated shape of the fish, whereas the classifier *wei* paints the picture of a fish with a large flowy tail.

However, which classifier should be used is not always an easy call for a non-speaker. For example, complexity arises when multiple features of the noun lead to diverging classifiers:

[1] In most Chinese dialects, classifier usage is also present. The specific rules about when to use which classifier varies considerably across dialects (Tai, 1994). In some northern dialects, classifiers are occasionally omitted in the spoken language.

Paper has a flat surface and is a thin slice. The correct classifier usage for *paper* is *zhang,* a classifier for things with a flat surface (other nouns in this category: table, bed, ticket), rather than *pian*, a classifier for thin slices of objects (other nouns in this category: bread slice, meat slice, leaf). This is also an example where the resemblance between some members sharing the same classifier is not intuitive to a non-speaker. Arbitrariness may also arise when the language has evolved and the apparent resemblance between nouns has faded: the classifier for *lamp* and *tea* are both *zhan* because the shape of a lamp used to resemble a teacup (but not anymore). Moreover, for abstract nouns, the resemblances in perceptual properties are even less apparent. In sum, similar to grammatical gender, the classifiers' semantic relation with their associated nouns can be apparent at times but ambiguous at other times. Consequently, the appropriate usage of a classifier is considered one of the hardest parts to learn in Chinese.

Despite the complexity and ambiguity at times, some argue that classifiers are rarely if ever arbitrary grammatical rules but rather meaningful categorization markers (Tai, 1994; Zhang et. al, 2011). Nouns that share a classifier, despite having no apparent connections due to the reasons above, would still be considered more similar with each other by a native speaker (Zhang & Schmitt, 1998; Imai & Saalbach, 2010). This can be explained by the prototype theory of categorization where central members of the category bear the key features that other peripheral members of the category more or less resemble, while members of a category may be associated with one another in a 'family resemblance' style (Wittgenstein, 2010).

Schmitt and Zhang (1998) investigated the effect of classifier language on consumers' product choices with a series of experiments. They found that objects sharing a classifier were considered more similar to each other and tended to be recalled together in a memory task, indicating that the classifier was used to mentally organize concepts. Chinese participants also

were more likely than English speakers to first mention the classifier-related common features when asked to give a list of common features between two objects (Schmitt and Zhang, 1998). Moreover, the classifier effect was demonstrated in a lab-simulated consumer choice context where consumers drew inferences about certain product features based on the classifiers assigned by experimenters, which had downstream influences on their product choice (Schmitt & Zhang, 1998). Another study looked at participants' evaluation of advertisements and found that ads with a congruent representation of the product and its classifier were evaluated more positively (Zhang & Schmitt, 1998). In sum, quite like how the grammatical gender of an otherwise inanimate and genderless object may give it gender-specific characteristics, the different classifiers may carry with them the particular features of their own groups and may lead people to project these features onto the concept that they are modifying.

However, note that the findings in the classifier effect are not always so strong and clear-cut. Imai and her team, for example, demonstrated a similar classifier effect when comparing Chinese speakers with Japanese and German speakers but cautioned that the observed cross-cultural difference in the tendency to make classifier-based concept organizations was minuscule compared with the competing organizing schemes such as those by taxonomic and thematic relations (Saalbach & Imai, 2007). A more recent study comparing Chinese and English speakers' memory of classifier-sharing nouns when they were imbedded in meaningful sentences (as opposed to just words in Zhang & Schmitt 1998) found that Chinese speakers use classifier-based 'clustering' to aid their recall only when the classifier group was 'well-defined' (as opposed to arbitrary) (Gao & Malt, 2009, experiment 5). There is also evidence suggesting that the classifier effect was task-specific: more robust in similarity judgement tasks and general classification tasks (E.g. which objects are more similar to each other/belong to the same

category?) than in property extension tasks (E.g. which objects are more likely to carry the same kind of bacteria?) (Imai, Saalbach, & Stern, 2010). There is also evidence showing that classifiers only influence cognition when they are overtly present in a sentence (Gao & Malt, 2009; Huettig et al. 2010). For example, Huettig et al. found using eye-tracking devices that Chinese speakers look at the object sharing the same classifier as the one they hear simultaneously in their headphones, but only when the classifier is explicitly present in the spoken sentence (2010).

In summary, existing evidence suggests that classifier use can affect cognitive processing of the speaker, but the size of the effect is likely dependent on the particular classifier chosen (well defined or arbitrary), the presence of other competing organizational schema (e.g. taxinomic and thematic relations), and how much explicit attention was drawn to the classifiers themselves (e.g. whether the classifier was explicitly used/repeated in the study stimuli). As will be seen, the present studies provide evidence suggestive of a classifier effect in spontaneous use outside a laboratory setting (Study 3) as well as evidence of boundary conditions (in an experiment, the classifier required a prime to create assimilation and contrast effects).

1.4 Effect of Language in High-stakes Situations

Compared to the lab-simulated evaluation and memorization tasks used in the past research quoted above to demonstrate the influence of grammatical gender and classifier, predicting future stock performance is a much more realistic and high-stakes scenario. Would linguistic features bias perception and decisions in such a high-stakes situation where, supposedly, more care and deliberation would be put in? Past research has some indirect answers

where researchers have looked at how other important real-world life choices are influenced by linguistic features.

For example, speakers of languages without a marked future tense are more willing to sacrifice their present happiness for future well-being, engaging in more forward-thinking behaviors (Chen, 2013). They would choose to save more for retirement, be more likely to use condoms, have more regard for long-term health by exercising more and smoking less, etc.[2] It seems that lacking the vocabulary to talk about the future as if it is separate and distinct from the present enhances the connection between one's current self and the future self, making the sacrifice for the future easier. Moreover, a correlational study suggests that among US immigrants, those with a grammatically gendered mother tongue are more likely than speakers of non-gendered languages to allocate household tasks according to gender identity norms (Hicks, Santacreu-Vasut, & Shoham, 2014). Worldwide, the more gendered the language of a country, the less gender equality in the country, as indicated by various indices (Prewitt-Freilino, Caswell, & Laakso, 2011), though this study similarly suffered from *Galton's Problem*. Although the causal direction is hard to establish, these studies provide preliminary evidence that grammatical features of a language reflect and reinforce certain aspects of a culture.

The financial market is undoubtedly high-stakes while at the same time bias-prone. This complex and perplexing man-made system embodies two contradicting elements: rationality and irrationality. On the one hand, the expectation is that meticulous thought and research must go into stock market investment decisions since real money is at stake. On the other hand, the

[2] The cross-cultural correlational studies like this would inevitably run into *Galton's Problem*. That is, cultures cannot be assumed to be independent of each other, so the co-occurring traits (language and future benefitting behaviors, for example) from a group of cultures could be due to the cultures sharing a common ancestor, rather than due to an independently-operating causal relationship between the two traits. A follow-up study used several methods to control for the dependency between cultures (Roberts, Winter, & Chen, 2015) and concluded that the correlation remained robust, albeit weaker, under most of the tests, except for the most stringent one. The recommendation was that experiments or case studies are needed to make the causal inference.

frenzied and reoccurring market bubbles blown up by hot air seem to indicate just the opposite of cold rationality, especially in hindsight. Hence, it may be a surprising but not far-fetched finding if grammatical features like classifiers and grammatical gender could influence high-stakes financial decisions, making the stock market an ideal environment to form an interesting yet stringent test of language effects.

It has always been the million-dollar question (quite literally) to predict where the market will go tomorrow. Some say that even Isaac Newton was daunted by the task and muttered that he could 'calculate the motions of the heavenly bodies, but not the madness of the people.' (Holodny, 2017) But its fickle nature does not stop investors from diving into the financial market and seeking opportunities to beat the market. Popular strategies include fundamental analysis that evaluates macro-economic factors and the companies' balance sheets to look for undervalued companies (e.g.: Abarbanell & Bushee, 1997), and technical (or chartist) analysis that studies the trends of past prices to predict movements (e.g.: Murphy, 1999).

We argue that investors are also influenced by a third latent "strategy," consciously or not. Rather than strictly sticking with the numbers on the balance sheets or the stock chart, they think of the stocks in terms of human traits such as agency, strength, steadiness, and emotion, etc.

Ben Graham famously described the market as the "temperamental" business partner, "Mr. Market," who is ridden with boundless optimism at one moment and bottomless pessimism at another (Graham, 1965). These personifying characterizations may be about the *people* making decisions and causing the market movements (e.g. the *people* are panicked, confident, etc.), or it may be that investors imbue the stocks, the companies behind them, or the market

system with human-like qualities, projecting social schemas on to them. Perhaps both tendencies occur.

Past research has analyzed the media portrayal of the market and found evidence of this personification and its relationship with the price trend. For example, Morris and colleagues surveyed end-of-day CNBC market commentary and found that commentators were particularly willing to use agent metaphors (as opposed to object metaphors or no metaphor) to describe the market's movement when the trend was upward and steady. Examples of the agent metaphor were "(6/3.1980) Financial Federation *leaped* …" and "(6/2/1980) ERC Corp. stock *soared*." Furthermore, when a stock was described using an agent metaphor, research participants were more likely to expect that the current trend of the stock would continue (2007).

Another study found that agent metaphors led to more optimism on the stock's outlook (Nicaise, 2014). In the study, participants saw neutral price charts and some excerpts of made-up stock performance evaluations with a slightly optimistic tone, containing different types of metaphors. The researcher manipulated the types of metaphor to range from emphasizing complete agency (the market is mobile, like a rational and responsible human being who knows where to go), to less agency (a health metaphor), and finally to no agency (a weather metaphor). The mobility metaphor that implied agency was the most effective in getting participants to invest.

1.5 Agency: the mechanism through which linguistic features affect stock performance prediction?

Agent metaphors imply that stocks possess the qualities of an agent. Specifically, an agent metaphor implies a sense of willful control over the self (decisiveness, rationality) and effective action in the situation (strength, efficacy). Both are stereotypically male traits, while

females are considered to be on the opposite pole, lacking emotional control, indecisive, weak, and passive. Connecting the dots, we hypothesize that the grammatical gender of the stock casts gendered expectations on the stock's agency, and consequently influences the stock's future direction. Specifically, many names of the stocks that are traded in Spanish-speaking countries could be coded as either grammatically masculine or feminine (hereby referred to as male or female stocks). We hypothesize that investors, knowingly or not, would expect male stocks to have more agency than female stocks.

Studies reviewed thus far suggest that imbuing "agency" to stocks leads to two expectations that may or may not converge: 1) continuation of the current trend (Morris et al., 2007, Study 1) – also called *momentum*[3] - or 2) going upward in general (Morris et al, 2007, studies 2-4; Nicaise, 2014). When stocks have gone up in the past, both 1) and 2) suggest that the more 'agentic' ones will be more likely to continue to go up in the future. However, when stocks have gone down in the past, 1) and 2) suggest opposite trends. The *agency = continuation* story would suggest that agentic stocks went up because going up is part of who they are (i.e. enduring characteristic of an agent) so they will more likely continue to do so compared to the less agentic stocks which likely went up by chance. Similar logics apply to going down. On the other hand, the *agency = going up* story could go like this: the more agentic stocks "want" to go up since it is

[3] An abundance of literature has examined the momentum effect and has attempted to explain the reason for its existence (Jegadeesh & Titman, 1993; Jegadeesh & Titman, 2001; Rouwenhorst, 1998; Moskowitz & Grinblatt, 1999; Menkhoff, Sarno, Schmeling, & Schrimpf. 2012; Asness, Moskowitz, & Pedersen, 2013). It has now been well established that the momentum effect consistently exists across time and countries (except in Japan and China; see Chui, Titman, & Wei, 2000). However, researchers have not yet reached a consensus on its cause. Popular theories explaining the existence of momentum fall into two categories: risk-based and behavioral. The behavioral models explain momentum as either an under-reaction or a delayed over-reaction (e.g. Daniel, Hirshleifer, and Subrahmanyam (1998), Frazzini (2006)), which could be caused by delayed information dissemination, self-enhancement thinking, lay theories of change, etc. The risk-based explanations try to find unique risks accompanying the high-momentum stocks to justify their momentum return. Furthermore, multiple factors could work together to result in momentum. Of course, should our momentum hypothesis turn out to be correct, neither of the grammatical features could completely account for momentum. because momentum in stock prices is found in cultures that do not have grammatical gender and do not use classifiers. The hypothesis is not that language effects explain all momentum phenomena, but that it may add to momentum effects in cultures where languages do have these features.

generally the more “desired” way to go. So when the past trend was up, the predictions go the same as the *agency = continuation* hypothesis. But the noteworthy part is that when agentic stocks had gone down in the past, they should “want to” and be able to reverse course and go back up. The less agentic stocks do not have equally strong efficacy so when they had slumped, they will slump further. In sum, hypotheses 1) and 2) predict similar results when the past trend had been upward but different results when the past trend had been down.

The agency story can also be applied to the classifier-stock context. Specifically, in Mandarin Chinese, a stock can both be described by an animal classifier 只 (zhi) and an object classifier 支 (zhi) - both are widely accepted as the appropriate usage. We hypothesize that describing a stock with an animal classifier would imply more agency. Further, priming stocks by having participants read sentences about animals should lead participants to expect greater agency when the stock is paired with an animal classifier.

As we will also see, however, agency is not a unitary construct. The way males have stereotypically greater agency than females (due to their strength or rationality) is different than the way animals have greater agency than objects (due to animals having a will and being capable of independent movement, rather than being strictly subject to the laws of inertia, as physical objects are). As will be seen, this leads to different patterns in the Spanish grammatical gender study contrasting male vs. female stocks and the Chinese classifier study contrasting animal vs. object classifiers.

1.6 The Present Research

The current research package aims to connect the dots reviewed above to examine the influence of language on the predicted future movement direction of stocks. Specifically, it shows that certain arbitrary-seeming linguistic features of the language can carry implications of agency, which in turn affect how speakers of that language perceive and predict stock movement directions. Namely, both being grammatically male (Spanish) and being mentioned with an animal classifier (Chinese) are associated with a higher perceived agency which leads to more optimistic forecasts of the future (i.e. predicting that the stock will go up).

In studies 1-2, grammatical gender in Spanish was examined as the focal linguistic feature, and stock markets of Spanish-speaking countries were analyzed. Both an archival study (study 1) and an experiment (study 2) demonstrated grammatical gender's effect on stock movement forecasts. Study 2 also attempted to determine the underlying mechanism for the gender effect by testing different aspects of agency: rationality and strength. In studies 3-4, classifier usage in Chinese was examined in the Chinese stock market with Mandarin-speaking participants. Study 3 was an archival study where Chinese stock commentaries were coded for their choice of the Zhi classifier and their prediction for the stock. Study 4 tested the same classifier effect with an experiment to try to establish the causal link.

CHAPTER 2: SPANISH GRAMMATICAL GENDER STUDIES

The following two studies demonstrate that seemingly arbitrary linguistic factors, such as the grammatical gender of the company's name, can influence people's expectations of its stock's future price trends.

2.1 Study 1: an Archival Study

Study 1 is an archival study using historical price data from stock markets in Spanish-speaking countries. We hypothesized that if investors who speak a grammatically gendered language indeed carry their gender stereotypes over to the arbitrarily assigned grammatical gender of a company, they should perceive a company with a masculine name to have more agency. Specifically, the male companies will be considered to have more strength and be less labile than female companies. This imagery of the company should be reflected in expectations for the future price trend of the company's stock.

As noted above, there are two competing hypotheses – '*agency = continuation'* and '*agency = up'* – which may suggest a clearer trend following past upward movement (i.e. agentic companies should go up more in the future compared to less agentic ones) than following past downward movement. Specifically, the 'agency = continuation' hypothesis (H1) suggests that seeing a company as strong, determined, and rational should lead investors to be more likely to predict a continuation of the current trend (i.e. large momentum), whereas seeing a company as weak, emotional, and capricious will lead to the expectation that the current trend won't necessarily persist (i.e. small or negative momentum).

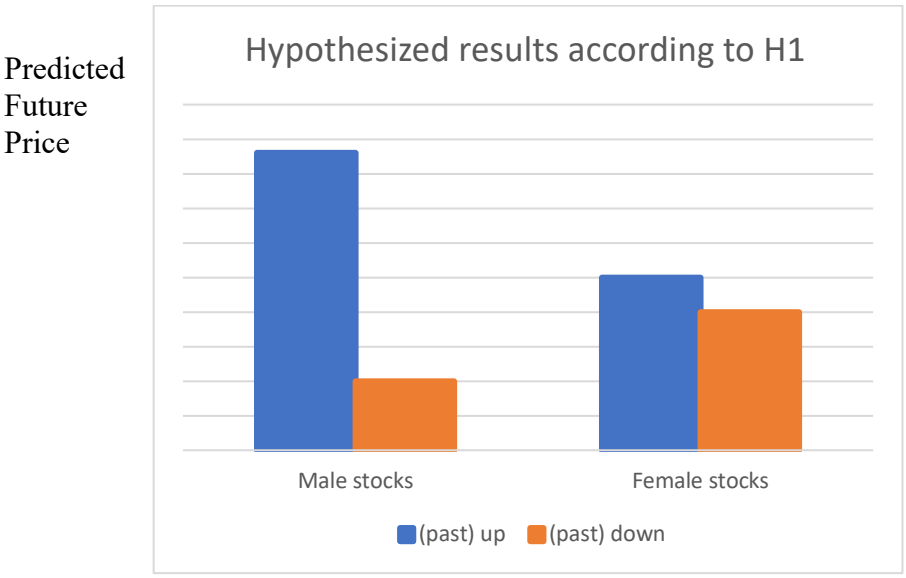


Figure 1. *Hypothesized results according to H1, study 1*

The 'agency = up' hypothesis (H2) suggests that more agentic companies should be able to go up more than less agentic ones regardless of the past trend.

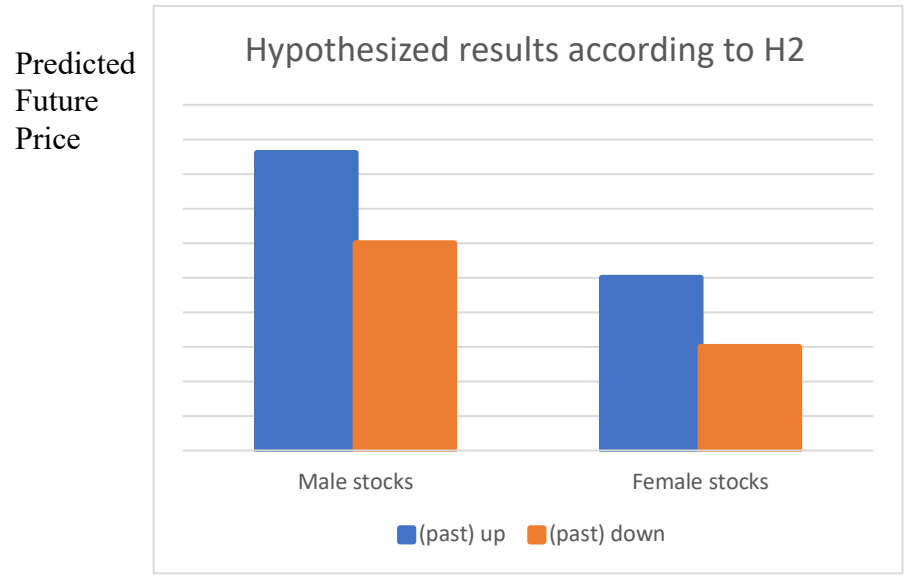


Figure 2. *Hypothesized results according to H2, study 1*

These expectations of investors will in turn translate into their investment behavior, and ultimately lead to the actual movement of the stock price. For example, when predicting that the current upward trend will persist, investors will buy the stock, thus indeed driving the price up.

Should the effect exist, it will most likely be strongest when most of its investors are native Spanish speakers. Therefore, we decided to focus on shares of Spanish-speaking countries' companies that are both incorporated and traded in Spanish-speaking countries. Companies that trade in stock exchanges located in non-Spanish-speaking countries are more likely to have a larger proportion of traders that don't speak Spanish and are therefore not included in this sample.

2.1.1 Method

Data collection. A list of all companies that are originated and traded within all Spanish-speaking countries (CIA factbook, 2015) was obtained from the Bloomberg database[4] (*n* = 1041). Countries were considered Spanish speaking if Spanish is the official language and a majority language (if a percentage is given). Three native Spanish-speaking research assistants coded the grammatical gender of the company names into three categories: *feminine*, *masculine*, and *ambiguous/neutral*. The third category was a catch-all category for the names that cannot be categorized as feminine or masculine[5]. Cohen's Kappa for inter-coder reliability between each pair of the coders (excluding the ambiguous/neutral category) was approximately 0.8.[6] We disregarded names with disagreement or names categorized as ambiguous/neutral. A fourth person – a South American native fluent in Spanish – reviewed the work of the three coders and

[4] The Spanish speaking countries include Argentina, Bolivia, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Uruguay, Venezuela, and Spain.

[5] Most of these names are foreign words, made-up words, or abbreviations. The neutral category looked similar to the female category and was significantly below the male category for overall and upward momentum (overall-day p=.005; overall-month p=.031; buy-day p=.001; buy-month p=.004). For downward momentum, neutral did not differ significantly from the male category for either day or month computations (both ps≥.42). The neutral category did not differ from the female category for any momentum measure (all ps≥.20). For descriptive statistics for the neutral category, please see Appendix VI).

[6] If the ambiguous category is included, Cohen's kappa for each pair was approximately = .6. This makes sense since having an ambiguous grammatical gender by definition means that coders will tend to disagree on the gender. See Appendix A for some examples of masculine, feminine, and gender ambiguous/neutral company names.

trimmed the list further. The remaining companies were: $n_{male} = 234$, $n_{female} = 313$. Basic financial information for these companies such as their price to earnings ratio (P/E) and book value ratio to market value (BV/MC) was collected for this set of 547 companies.

The historical end-of-day price data from 2000 to 2015 from all actively trading companies from Spanish-speaking countries were gathered from Bloomberg[7]. To avoid the calculated momentum index being swayed too much by any single day's abnormality, we removed any company with fewer than 50 days' price data. The final sample averaged around 350 companies, depending on the variable and calculation strategy employed.

Momentum Calculations. To quantify trend continuance (H1) and upward movement (H2), we calculated the rate of return obtained from buying past rising stocks and short-selling past falling stocks (i.e., invest using the momentum strategy) averaged across multiple rounds of operations (defined below). An average return was calculated separately for past upward (i.e., $momentum_{up}$) and past downward (i.e., $momentum_{down}$) trends to gauge H1 and H2 separately. An additional rationale for examining $momentum_{up}$ and $momentum_{down}$ separately was that since the market historically moves upward over the very long run due to growth of the economy and inflation, the momentum strategy for selling past loser stocks when defining "loser" in absolute terms (a price decrease) may not be profitable and the return of $momentum_{down}$ may turn out to be negative. Indeed, past research in momentum has found that longing strategies (buying past winners) are more profitable than shorting strategies (selling past losers) and that the return from

[7] Many companies have missing data points due to the stock shares not trading on a particular date. Across all companies in our sample, the number of available days per company ranges from 0 to 3959, and only nine companies have complete data for every day.

momentum was primarily driven by the longs (for example, see Jegadeesh & Titman, 1993; Asness et. al. 2014)[8].

To calculate momentum for each stock, we observed stocks for a certain period (T1), (hypothetically) bought them if their price increased during T1 and short-sold[9] them if their price decreased. This position was held for another period (T2). In both scenarios the outcome variable was calculated as [(Price at end of T2 – Price at beginning of T2)/ price at beginning of T2] which indicates by how much the stock had increased its price during T2. If price increased in T1, then a positive outcome variable indicates upward momentum ($momentum_{up}$); if price decreased in T1, then a negative outcome variable indicates downward "momentum" (downward movement) ($momentum_{down}$). In general, a positive outcome variable indicates price appreciation over time. Operations were carried out every day until all data from the company were used. We also calculated the outcome monthly on the first trading date of the month, following the convention in the finance literature. We present the results from both the monthly and the daily approaches below[10].

Regarding the lengths of T1 and T2, there is no consensus in the momentum literature, so we included all three "optimal lengths" suggested by different researchers (Jegadeesh & Titman,

[8] There are alternative ways of defining the winner and loser stocks. For example, stocks can be compared with all the other stocks in the same portfolio, forming a rank based on their performance during T_1. The relative winner - loser status can be defined as the first and last X% of the stocks. Researchers in the past have chosen X=10% or X=30%, based on the size of their portfolio. In this way, the investment decision is based on the relative performance of the stock during T_1, rather than its absolute performance. Thus, hypothetically, in a bull market the loser stocks can also be rising, only slower than the winners. Some have argued that the relative winner-loser status more accurately defines momentum than absolute gains and losses (Asness, Frazzini, Israel, & Moskowitz, 2014). Another way to define winner-loser status is to use the slope of the best fitted regression line between time and price during T_1. This definition considers the added information from all data points between the starting and ending points of T_1. However, it is unclear which definition is more similar to the actual thought process of investors when evaluating momentum. We thought the simplest strategy (buy stocks that have gone up, short those that have gone down) was the best in this regard. Future work can explore how the other definitions of momentums affect the magnitude of the gender effects.

[9] Short-sell refers to the operation that benefits from a falling stock price. Stocks are "borrowed" at the beginning of the period and sold at the price P_1. At the end of the period stocks are bought back at the price P_2 and "returned". The investor makes a profit if the stock price fell during the period, i.e. $P_2 < P_1$.

[10] In reality, operations often are not carried out so frequent as once per day due to trading costs, bid-ask spreads, etc., which partially explains why simulations often use monthly rather than daily intervals.

1993, 2000; Chui, Titman, & Wei, 2000; Goyal & Wahal, 2013). A consistent gender difference across different holding and trading periods would be an indication of the robustness of the result. Moreover, some studies have suggested that the momentum effect is only prominent in the medium run, while in the very short run as well as in the long run a reversal effect tends to emerge where buying past losers and selling past winners (exactly the opposite of the momentum strategy) becomes profitable (Jegadeesh, 1990; Lehmann, 1990). Consequently, some researchers advocate for excluding the very short run from analysis to acquire a larger momentum effect (Chui, Titman, & Wei, 2000; Goyal & Wahal, 2013). This exclusion is achieved by inserting a waiting period (T_{wait}) of 1 month between T1 and T2. We employed three different timing strategies in total: 1) observe whether a stock has risen or fallen over 6 months, skip a month, measure its performance over the next 6 months (6-1-6 strategy); 2) observe for 11 months, skip a month, measure performance over the next 3 months (11-1-3 strategy); and 3) observe for 12 months, measure performance over the next 3 months (12-0-3 strategy)[11] (Jegadeesh & Titman, 1993, 2000; Chui et al., 2000; Griffin, 2003).

Over the 15 years timespan, a stock could be bought or sold many times depending on how it performed during the observation period. All but approximately a dozen stocks (depending on the use of monthly or daily calculations) thus had a score for $momentum_{up}$ (the average outcome after the stock had gone up during the observation period) and $momentum_{down}$ (the average outcome after stocks had gone down during the observation period). Finally, an overall momentum value was calculated by subtracting $momentum_{down}$ from $momentum_{up}$

[11] The seminal paper by Jegadeesh & Titman (1993) examined a number of timeframes, finding that the 12-0-3 strategy was most profitable. They also extensively analyzed a 6-month observation, 6-month holding period, reporting that results for this strategy "are representative of the results for the other strategies" (p. 69) (see also Jegadeesh & Titman, 2001). The skipping of one month attempts to take advantage of the well-known 1-month reversal effect (Jegadeesh, 1990).

(weighted by how many times the stock was longed vs. shorted). Essentially it is measuring how much more do the stock tend to rise following previous uptrends than following previous downtrends. It is denoted by $momentum_{overall}$. We had a primary coder build the hypothetical trading algorithm for processing the 15 years of data. As a check on accuracy, we also had a second coder construct an algorithm and process the data independently. These results were extremely similar, and the average result is presented below.

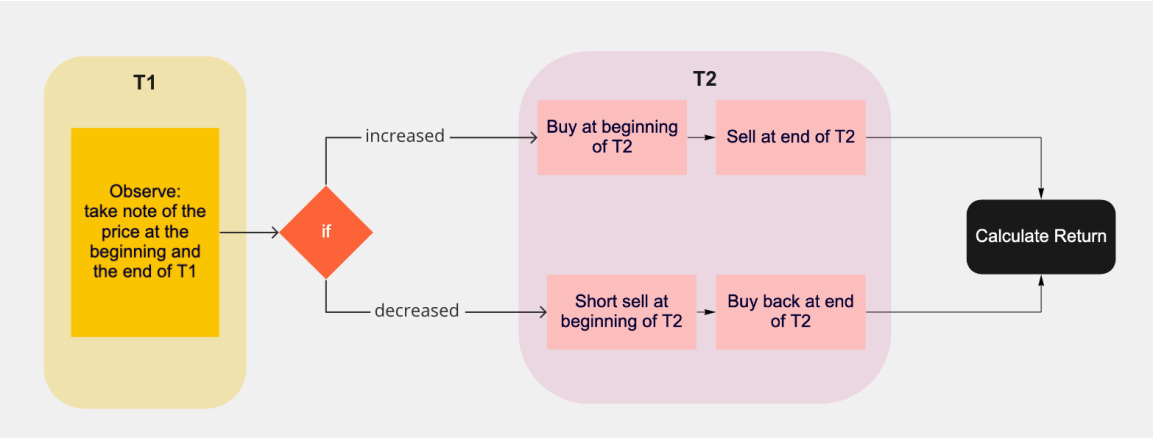


Figure 3. *study 1 procedure*

Histograms and boxplots of the momentums showed that some companies' stock momentums were potential outliers. Some outliers had such extreme values that they were more than 10 times larger than the average. We expect that grammatical gender is irrelevant for extreme cases, which are likely due to major events (for example, a company heading toward bankruptcy or patenting a drug that rockets its profitability). Therefore, we used Tukey's "inner fences" – 1.5 times the Inter-Quartile Range above the 75th percentile and below the 25th percentile – as the criteria to determine and remove outliers[12].

[12] Conclusions *do* hold if nonparametric methods are used for analysis with the whole sample. We examined Kendall's tau of grammatical gender (dummy coded) and each of the three momentum outcomes (overall, up, and down). The male-female dummy is significantly correlated with $momentum_{overall}$ by month or by day (overall monthly p=.029, overall daily p=.014) and with $momentum_{up}$ by day (p=.028); It is marginally significantly correlated with $momentum_{up}$ monthly (p=.061); the downward momentum is not (monthly p=.27, daily p=.18).

Hypotheses revisited. We hypothesized that $momentum_{overall}$ and $momentum_{up}$ should be positive, consistent with the literature. Also, a clear gender effect should emerge for both, such that masculine-named companies will show a higher momentum than feminine-named companies. Masculinity is associated with being strong, determined, and levelheaded. Once a goal is set, it is hard to make the person yield. Mapping this gender stereotype onto the movement of stock prices would result in higher upward momentum. On the other hand, femininity is associated with having more emotionality, weakness, and less predictability. Applying this gender stereotype to the investment scenario could indicate that previous price trends would not necessarily signal future trends, so investors might expect smaller upward momentum from stocks with feminine names.

With regard to $momentum_{down}$, depending on how much the overall market displayed an upward general trend (i.e., a "bull" market), it was unclear whether shorting past losers would be profitable or not. Therefore, $momentum_{down}$ could be positive or negative. More importantly, the two aspects of agency work in *opposite* directions when the past trend was falling. If agency implied continuity, then stocks going down should continue to go down. If agency implied upward movement, then downward stocks should reverse course and go up, although perhaps with a smaller degree of increase than $momentum_{up}$. Thus, the gender effect of $momentum_{down}$ was also unclear. Assuming that male stocks were seen as more agentic than female stocks, then when the stock has fallen in the past, the continuity aspect of agency would predict that male stocks would be more likely to keep falling than female stocks in the future, while the "going up" aspect of agency would suggest the reverse. If the two effects were both present and exactly cancel each other out, then there would be no gender effect for $momentum_{down}$.

2.1.2 Results

For ease of discussion, I will hereafter use "strategies" to refer to the combination of different observation and holding periods (6-1-6, 11-1-3, & 12-0-3) and calculation frequency (by day & by month). Averaging over the three holding-and-observation periods, male stocks generally exhibited higher overall momentum than female stocks – a result that held using both daily prices ($M_{day,male}$ = .031, $SD_{day,male}$ = .041 vs. $M_{day,female}$ = .018, $SD_{day,female}$ = .037; t(331) = 2.97, p_{day} = .003, *r(333)* = -.16) and monthly prices ($M_{month,male}$ = .031, $SD_{month,male}$ = .040 vs. $M_{month,female}$ = .020, $SD_{month,female}$ = .038, t_{month} (327) = 2.49, p_{month} = .013, r(329) = -.14) (see Table 1). The effect size was small, but perhaps larger than one might expect, given markets' tendencies to eliminate exploitable irrationalities.

Table 1. *$Momentum_{overall}$ Means for Male Stocks and Female Stocks*

Calculation Frequency	T_1-T_{wait}-T_2 (in months)	Gender of Stock	*N*	*M*	*SD*	*t*	*P*
Day	6-1-6	Male	144	.043	.055	2.17	.031
		Female	180	.029	.060		
	11-1-3	Male	149	.020	.035	0.98	.326
		Female	179	.017	.031		
	12-0-3	Male	148	.024	.033	2.01	.045
		Female	182	.017	.030		
	Average	Male	148	.031	.041	2.97	.003
		Female	185	.018	.037		
Month	6-1-6	Male	143	.042	.054	1.66	.097
		Female	175	.031	.059		
	11-1-3	Male	146	.024	.033	1.98	.049
		Female	180	.017	.032		
	12-0-3	Male	148	.027	.034	2.69	.008
		Female	183	.017	.031		
	Average	Male	146	.031	.040	2.49	.013
		Female	183	.020	.038		

$Momentum_{up}$ denotes the momentum of a set of stocks when they have gone up in the past. In other words, a larger value of $Momentum_{up}$ signifies that when the stock price rose in the

past, it is more likely to keep rising and/or have a larger increase. Large positive values in $Momentum_{up}$ indicates an upward movement following a previous upward trend and it signifies large upward momentum. Averaging over the three periods, we found that male stocks were more likely to rise during the holding period than the female stocks – a result that again held using either daily prices ($M_{day,male}$ = .059, $SD_{day,male}$ = .059 vs. $M_{day,female}$ = .036, $SD_{day,female}$ = .061; $t(321) = 3.40$, p_{day} = .001, $r(323)$ = -.19) or monthly prices ($M_{month,male}$ = .062, $SD_{month,male}$ = .063 vs. $M_{month,female}$ = .045, $SD_{month,female}$ = .070, $t_{month}(324) = 2.35$, p_{month} = .019, $r(326)$ = -.13).

Table 2. *$Momentum_{up}$ Means for Male Stocks and Female Stocks*

Calculation Frequency	T_1-T_{wait}-T_2 (in months)	Gender of Stock	*N*	*M*	*SD*	*t*	*P*
Day	6-1-6	Male	144	.087	.093	2.57	.011
		Female	176	.059	.101		
	11-1-3	Male	145	.042	.046	2.41	.017
		Female	176	.029	.050		
	12-0-3	Male	141	.044	.044	2.64	.009
		Female	181	.031	.048		
	Average	Male	144	.059	.059	3.40	.001
		Female	179	.036	.061		
Month	6-1-6	Male	145	.090	.100	1.88	.062
		Female	175	.068	.112		
	11-1-3	Male	145	.046	.043	1.71	.088
		Female	173	.037	.051		
	12-0-3	Male	144	.050	.047	3.21	.001
		Female	180	.033	.050		
	Average	Male	146	.062	.063	2.35	.019
		Female	180	.045	.070		

$Momentum_{down}$ stands for the movement of a set of stocks when they have gone down in the past. Based on our calculation formula, there is downward "momentum" (downward movement in price) when the value of $Momentum_{down}$ is negative. The larger the absolute value of $Momentum_{down}$ while being negative, the larger the downward movement is. On the other

hand, positive $Momentum_{down}$ indicates a lack of downward "momentum" (downward movement) and, instead, a tendency to reverse and go up. A larger positive $Momentum_{down}$ represents more upward movement following a previous downward trend. Our results showed positive $Momentum_{down}$ across different strategies, with a smaller magnitude than $Momentum_{up}$ (Table 3). This suggests that stocks tended to reverse back to upward moving rather than continuing their downward trend. The rise in male stocks was not significantly different from the rise in female stocks when calculations were made daily ($t(329) = 1.29$, $p = .200$) or monthly ($t(320) = 1.51$, $p = .133$).

Table 3. *$Momentum_{down}$ Means for Male Stocks and Female Stocks*

Calculation Frequency	T_1-T_{wait}-T_2 (in months)	Gender of Stock	*N*	*M*	*SD*	*t*	*P*
Day	6-1-6	Male	141	.046	.077	.14	.892
		Female	176	.045	.087		
	11-1-3	Male	140	.023	.044	1.18	.240
		Female	171	.017	.043		
	12-0-3	Male	142	.021	.046	1.04	.300
		Female	172	.016	.041		
	Average	Male	148	.038	.063	1.29	.200
		Female	183	.029	.062		
Month	6-1-6	Male	146	.058	.092	1.16	.248
		Female	174	.046	.088		
	11-1-3	Male	139	.025	.044	1.40	.163
		Female	173	.018	.046		
	12-0-3	Male	143	.024	.048	1.11	.266
		Female	176	.018	.044		
	Average	Male	144	.039	.058	1.51	.133
		Female	178	.030	.058		

Further analyses. That male stocks significantly outperform female stocks following an upward trend and tended (nonsignificantly) to also do so following a downward trend suggests a very surprising possibility: that male stocks (vs. female) would more likely end up being overvalued relative to their fundamental values, at least during the momentum holding period

before any mean reversion. Such overvaluation would be surprising because it would be a very easily exploitable margin of profit that should quickly be exploited away (for example, by people buying female stocks at a discount and selling male stocks at a premium). To scrutinize this possibility, we looked for the gender effect in other conventional measures of overvaluation: the price/earning ratio (P/E) and the book value/market capitalization ratio (BV/MC). P/E indicates how much the market (i.e., investors in aggregation) is willing to pay for one share of a stock today based on its past 12 months' earnings, while BV/MC denotes how much the company's net assets weigh in relation to each dollar that the investors are paying for in the stock market. Both ratios help assess the worthiness of the stock in the eyes of investors. Hence, larger P/E and smaller BV/MC ratios in male stocks than female stocks would show that investors value male stocks over female stocks.

Data was gathered from Bloomberg[13] and the Emerging Markets Information Service (EMIS). Price-earnings ratios that were negative were discarded because negative values have no conventional interpretation. Negative BV/MC (due to negative bookvalue) was kept in. Data from the two sources were averaged, and then the two metrics were also averaged after the standardization of both, and the reverse coding of BV/MC brought them onto the same scale. We then removed outliers beyond Tukey's "inner fences" (fences were calculated not including the gender neutral or ambiguous companies). Male companies (M = .012, SD = .140) are marginally significantly overvalued compared to female companies (M = -.016, SD = .137, $t(308) = 1.74$, $r(309) = -.10$, $p = .083$). However, unlike the findings about momentum, this effect was reliant on distributional assumptions. If one runs a nonparametric test between the male-female dummy

[13] Because many data points on P/E and BV/MC were missing (Bloomberg: $n_{P/E}$=259, $n_{P/Emissing}$=151; $n_{BV/MC}$=337, $n_{BV/MCmissing}$=73), we also gathered data from an alternative source: the Emerging Markets Information Service (EMIS). Combining data from the two sources recovered some of the missing data (Combined: $n_{P/E}$=304, $n_{BV/MC}$=353). For each company, input from the Bloomberg and EMIS database was averaged.

variable and the valuation metric on all the data, Kendall's tau is not significant ($r(347) = -.06$, $p = .187$). We find the data about male stocks being overvalued to be suggestive, but mixed, and needing further research to investigate its robustness.

We also gathered the stocks' sector information from the same source and examined whether the gender effect holds after controlling for sectors. There are ten sectors in our sample, five of which include an adequate number (> 10) of both male and female companies to be compared. Looking at just the subsample with the five sectors ($N = 270$), the main effects model ANOVA showed that even after controlling for sectors, the gender effect was mostly similar for overall momentum (by-day: $F(1,238) = 5.946$, $p = .015$; by-month: $F(1,236) = 1.899$, $p = .17$) and upward momentum (by-day: $F(1, 233) = 9.431$, $p = .002$; by-month: $F(1,234) = 5.303$, $p = .022$). For downward momentum, there was still no significant gender effect.

We also examined whether the gender effect differed based on sector using the full ANOVA model with interaction. There was one significant gender X sector interaction (upward momentum by-month; $F(1,230)=3.30$, $p=.012$) and two such marginally significant interactions (upward momentum by-day; $F(1,229)=2.26$, $p=.063$ and downward momentum by-month; $F(1,228)=2.398$, $p=.051$). For all three, the effect was similar: namely, male (vs. female) outperformance was primarily driven by the financial and consumer staples sectors. There were no significant or marginally significant gender X sector interactions for overall momentum.

2.1.3 Discussion

Tested with multiple observation periods and holding periods and with both calculation frequencies (by days and by months), $\text{momentum}_{\text{overall}}$ consistently differed based on the gender of the company name in Spanish. Decomposing the overall momentum revealed that the gender

difference resided primarily in $momentum_{up}$: male stocks were more likely to keep rising when they had risen in the past compared to female stocks. There was no momentum for downward trends, rather a tendency to reverse direction. The gender difference in $momentum_{down}$ was nonsignificant, although its trend together with the gender difference in the ups suggested possible overvaluation of the male stocks. However, analysis of standard value metrics only showed mixed evidence of overvaluation.

Looking back at the two competing hypotheses: both were supported by the strong gender effect of $momentum_{up}$, while the $momentum_{down}$ data favored (nonsignificantly) H2 (male = up). Specifically, archival data showed an unclear trend that following a past downward movement, male stocks – supposedly with their higher agency - were more likely to reverse back up than female stocks.

If the gender differences in $momentum_{up}$ and $momentum_{overall}$ are indeed a result of gender stereotypes carried over by the grammatical gender of the company names, one possible mechanism could be that stocks with masculine names are considered to be more agentic. There are two aspects of agency potentially relevant in this scenario: 1) male stocks could be considered strong-willed and determined like men, thus once a direction (especially an upward direction) was set, continuation was likely to follow; or 2) male stocks are seen as more rational and predictable like men, while female stocks are more emotional and irrational like a woman, resulting in the momentum expectations for them being different. Study 2 is an experiment that will test the agency hypothesis and also examine causality.

2.2. Study 2: an Experiment

This study hypothesized that the gender effect could work through either the strength route, or the rationality route, or both. That is, the reason for male companies' having higher momentum than female companies could either be that stereotypically 1) males (vs. females) are considered to be strong and will forcefully carry on with the direction that they are going, or that 2) males are considered to be rational, predictable, and level-headed and will thus show greater continuity (as compared to more irrational, emotional, unpredictable females). In this experiment, participants saw analyst commentary on male and female companies that nudged them to think in terms of either strength/weakness or rationality/emotionality. Also, to counteract the general tendency to assume a continuation of the trend, as shown by the existence of momentum related profits worldwide (except for Japan, see Chui et. al, 2000), news and analysis could provide participants with some reasons to predict a trend reversal, helping to avoid the maxing-out of momentum predictions. Our hypothesis was that analyst commentary that fit the gender stereotype should be especially persuasive, but whether it would be beliefs about forcefulness or rationality that do the work was an open question. Finally, we also measured participants' implicit and explicit gender stereotypes to see whether they were correlated with participants' tendency to predict future stock movement in gender-specific ways.

2.2.1 Method

One hundred and eighteen Spanish-speaking participants (Male: N=71; Female: N=47)[14] were recruited from Amazon Mechanical Turk online[15]. Participants completed a short screening

[14] Results were similar for male and female participants (for overall momentum, momentum up, momentum down, and overvaluation, all gender of participant X gender of stock interactions had Fs≤.44, all ps≥.51).
[15] Due to scarcity of native Spanish speaking participants in the M-Turk participant pool, data was collected in two batches such that participants from batch 2 were recruits from a previous Spanish-related study, who were originally excluded from batch 1.

of ethnicity and language at the beginning to select only those fluent in Spanish and at least partially of Hispanic origin. Each participant was then presented with information about eight different Latin American companies, including the name of the company, a piece of a recent news article about the company, a past stock performance chart, and some comments on the company's prospects based on the news and the chart from an expert market analyst (Appendix B).

The company information contents were not based on real-world events but were entirely created by the researchers, and the different contents within and across participants served as the experimental manipulation. Each participant read about eight different companies and was asked to predict the company's stock performance in the future. For each company, they were given the name, a graph showing its historical stock trend, a short news excerpt allegedly from Yahoo Finance, as well as a piece of expert commentary. For each participant, half of the company names were masculine, and half were feminine; half of the past trend graphs went upward, and the other half went downward. Each company was presented only once to each participant. The events reported in the news article were ambiguous and could be interpreted either positively or negatively: half of the news articles could have either a high rationality interpretation or a low strength interpretation, while the other half could reflect either high strength or low rationality. Lastly, the experts' commentaries catered to one of the possible interpretations of the news, and consequently adopted either a negative or a positive tone. All variation combinations were counterbalanced. In summary, this study employed a 2 (male stock – female stock) X 2 (up – down) X 2 (analyst comment positive or negative) X 2 (analyst comment on forcefulness vs. rationality) factorial design. The first two factors were within person factors, while the rest

Analyses were done on the collapsed data set (n=118) to increase statistical power. The same analyses were also done on each batch of participants separately and results are reported in footnote 16.

varied partially within and partially between participants (For example, news A may be interpreted as high forcefulness for person one but low rationality for person two; news B may be interpreted as low forcefulness for person one but high rationality for person two). The parts that were between participants were counterbalanced.

Figure 4. *An example of the company information presented to participants with the analyst in the first panel framing the news in terms of strength and the analyst in the second panel framing it in terms of irrationality*

Panel 1

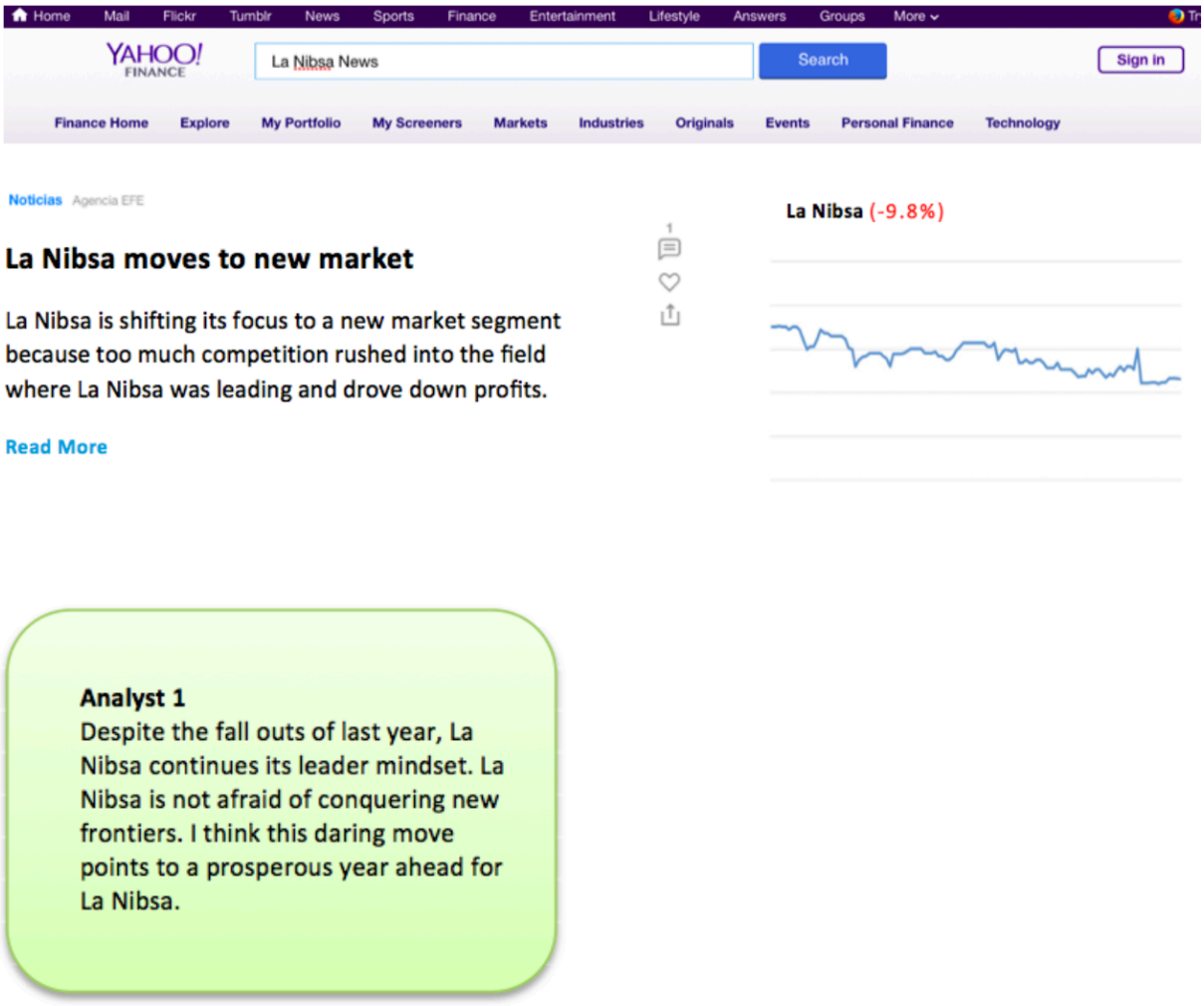
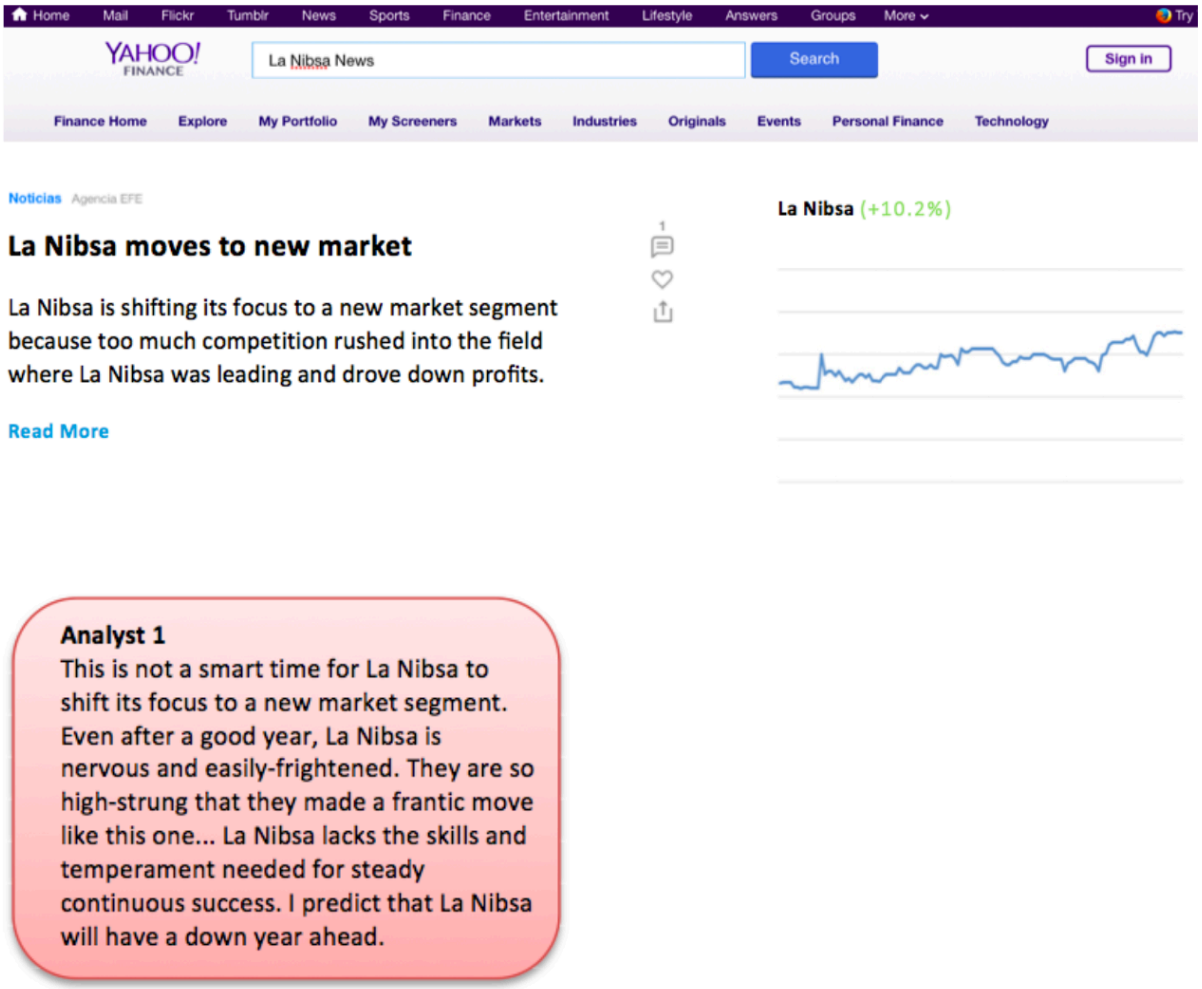


Figure 4 (cont.) Panel 2



After reading each companies' information, participants first answered what the experts thought about the company on a 100-point scale. Since the experts made their opinions very clear at the end of their commentaries (e.g. "I predict that XXX will have a down year ahead"), this question served as a quality check. Then, participants chose whether to buy or sell the company's stock based on all given information in a binary choice question and indicated their confidence about the decision in the question below on a 10-point scale. Deciding to buy after an upward past trend indicates that participants expected upward momentum; deciding to sell after a

downward past trend indicates an expectation of downward momentum. On the flip side, buying after a downward trend and selling after an upward trend would indicate an expectation of reversal. The confidence measure is treated as the measure of likelihood of taking action. $Momentum_{up}$ simply averages all decisions following uptrends while $momentum_{down}$ averages all decisions following downward trends. $Momentum_{overall}$ is the difference between up and down because it shows how much *more* stocks are expected to go up after a previous rise vs. fall. To raise the stakes, participants were told that their performance would be linked to a bonus: each correct prediction would earn them 10 cents of bonus payment (20% of the base compensation), and each incorrect prediction would cost them 10 cents. In the end, everyone received 20 cents of bonus.

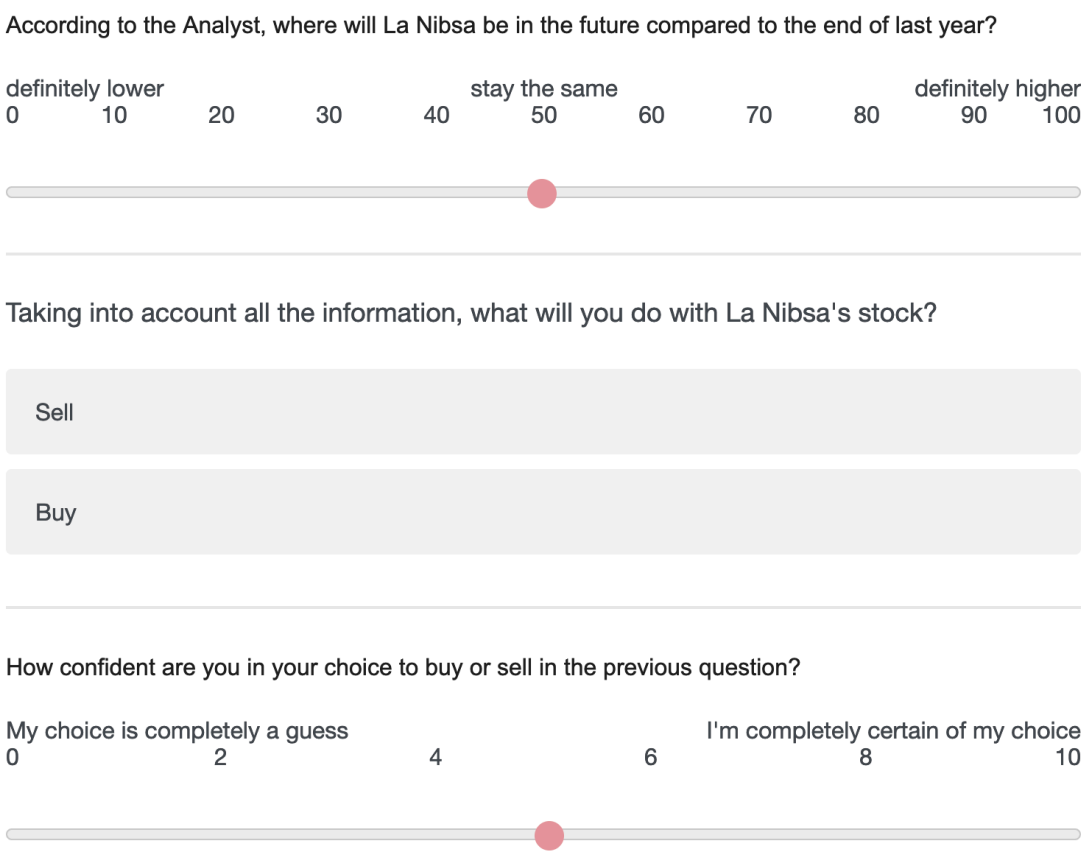


Figure 5. *Survey questions on stock performance prediction*

After completing the set of questions for all eight companies, participants went on to answer whether a list of traits applied more to males or females. This served as an explicit measure for how strongly they endorsed certain gender stereotypes. Five traits related to high stereotypical male rationality were rational, emotional (reverse-scored), good at math, predictable, and likely to make a big deal over unimportant matters (reverse scored). Five traits related to high stereotypical male forcefulness were determined, wishy-washy (reverse scored), obedient to authority (reverse scored), likely to make compromises (reverse scored), and

efficacious.[16] Participants then completed a memory test where they matched the names of companies to adjectives that were used by the expert analyst to describe the company in the previous task. The rate of correct and false memory was of interest. We hypothesized that participants would more readily remember the name-adjective pairs that were congruent with gender stereotypes (e.g. "La Nibsa – nervous and easily-frightened"). Moreover, the extent to which a participant misremembered in a stereotypical fashion would moderate the main effect of grammatical gender on perceived momentum. Additional demographic information was collected at the end of the study about participants' financial literacy, ethnicity, and Spanish fluency.

2.2.2 Results

The gender effect was constructed as a within-subject variable. For each participant's response regarding each of the eight companies, a prospective score was calculated by recoding the buy/sell response as buy = 1 and sell = -1 and then multiplying it by the confidence score[17]. The resulting prospective score ranged from -10 to 10. A higher prospective score indicated that the participant was more confident that the company's stock would rise in the future. Coupling with the company's past trend, prospective score for stocks that had been in an uptrend minus

[16] We created the composite variables *rationality* and *forcefulness* based on conceptual clarity, despite that not all items loaded cleanly as expected. Three other traits were stereotypically masculine but we did not a priori see them as related to the two factors: "tall," "likely to wear makeup" (reversed), and "good at relating to people" (reversed).

[17] The confidence score can be understood as a measure of the 'likelihood to act' if trading were done in the real world. The more confident one is about their buying/selling decision, the more likely they will actually participate in buying/selling the stock (Rieger, 2022), hence casting a real influence on the price of the stock. If the confidence measure is dropped from the analysis, conclusions on the key findings remain the same: male stocks are expected to have higher overall momentum and upward momentum than female stocks (up - down: $t(117) = 2.89$, $p = .005$; up: $t(117) = 3.40$, $p = .001$). This effect was again driven by negative comments that stuck more to female stocks (up_criticize: $t(116) = 3.43$, $p = .001$). Downward momentum showed no gender difference (down: $t(117) = -.90$, $p = .372$). There was also a trending male over-valuation evidence (collapsed across up and down: $t(117) = 1.74$, $p = .085$).
If the quality check filter is applied, all results above remain robust, except for the male over-valuation finding, which becomes significant ($t(99) = 2.20$, $p = .03$).

prospective score for stocks that had been in a downtrend would represent the overall expected momentum for that set of companies.

As will be recalled, we had a quality check such that respondents had to indicate whether the analyst had predicted the stock would rise or fall in his or her commentary. We present results for all respondents as well as for only respondents who answered our quality check questions at above a chance level (they got at least 5 of the 8 questions correct). This has implications for whether respondents were overvaluing male stocks in their predictions, which will be discussed later.

Momentum. Recall that positive value of momentum indicates an expectation of price appreciation over time. Overall, there was a significant difference in expected momentum when a company adopted a male name (M = 4.34, SD = 5.45) compared to a female name (M = 2.45, SD = 5.18). t (117) = 3.07, p = .003. The momentum for male companies was higher than the momentum for female companies. This finding is consistent with study 1 where $momentum_{overall}$ showed a significant gender effect for most of the strategies. If we exclude the 18 respondents in the present study who did not answer the quality check questions at above a chance level, male names are again expected to have a higher momentum (M = 4.01, SD = 5.11) compared to female names (M = 2.41, SD = 5.00). t (99) =2.45, p = .016.

Breaking down the momentum by past trend showed that the significant gender effect emerged only when the past trend was upward (male up momentum = 2.60, SD = 3.80 vs. female up momentum = 1.19, SD = 3.52, t(117) = 3.40, p = .001). As expected, participants predicted male stocks would show more upward momentum than female stocks. The downtrends conditions did not differ on average between companies with male names and female names (male down momentum = -1.74, SD = 4.00 vs. female down momentum = -1.27, SD = 3.87,

$t(117) = -1.09$, p = .278), though the trend was to believe that male companies' stocks would show more downward momentum than female companies would. The general conclusion is consistent with the findings from study 1 where a gender effect was prominent with $momentum_{up}$ but not with $momentum_{down}$. Excluding the participants in the present study who did not answer the quality check questions at above the chance level, this also remained true: the effect for upward-trending stocks was significant (male up momentum = 2.66, *SD* = 3.55 vs. female up momentum = 1.20, *SD* = 3.30, $t(99) = 3.37$, $p = .001$) and not significant for downward trending stocks (male down momentum = -1.35, *SD* = 3.95 vs. female down momentum = -1.21, *SD* = 3.74, $t(99) = -0.30$, p = .762).

With regard to the mixed findings that male stocks were marginally significantly overvalued in study 1, we collapsed over upward and downward trending stocks to see if in the experiment male stocks were expected to rise more than female stocks regardless of the past trend. Again, we found mixed results depending on whether the attention check criterion was applied. The full sample of respondents tended – not significantly – to expect male stocks to outperform female stocks (male movement = 0.44, *SD* = 2.80 vs. female movement = -0.03, *SD* = 2.64, $t(117) = 1.62$, $p = .109$). Excluding the participants who did not answer the quality check questions at above the chance level, this difference became significant (male movement = 0.66, *SD* = 2.75, female movement = -0.01, *SD* = 2.48, $t(99) = 2.16$, $p = .033$).[18]

[18] For the main tests, we also analyzed the two batches of data separately. In both batches, upward-moving male stocks were expected to have more momentum than upward-moving female stocks ($t(61) = 2.13$, $p = .037$; $t(55) = 2.69$, $p = .009$ in batches 1 and 2, respectively). In both batches, this was driven especially by upward moving stocks that analysts had criticized as low in rationality or forcefulness ($t(61) = 2.78$, $p = .007$; $t(54) = 2.60$, $p = .012$ for batch 1 and 2). For downward moving stocks, the male-female difference was not significant in either batch ($t(61) = .42$, $p = .68$; $t(55) = -1.16$, $p = .25$). Looking at overall momentum (up minus down), respondents in batch 2 expected male stocks to be higher ($t(55) = 2.66$, $p = .010$) while in batch 1 the trend was in the same direction but not significant, $t(61) = 1.73$, $p = .088$. Collapsing over both upward and downward moving stocks, respondents in both batches showed a nonsignificant tendency to expect male stocks to outperform female stocks $t(61) = 1.15$, $p = .26$; $t(55) = 1.13$, $p = .26$. Excluding those who did not pass the attention check, all the conclusions above remain the same, except for the male overvaluation which becomes significant for batch 1. See details below.
Excluding those who did not pass the attention check, in both batches, upward moving male stocks were expected to have more momentum than upward moving female stocks ($t(50) = 2.68$, $p = .010$; $t(48) = 2.09$, $p = .042$ in batch 1 and 2,

Further analyses. The expert message content tried to pinpoint the specific gender stereotype that was at play. Results showed that the effect was only present when the expert's commentary was negative, with both the low rationality (M_{male} = -1.60, SD = 6.62; M_{female} = -4.58, SD = 5.29; $t(59) = 3.14$, $p = .003$) and the low forcefulness commentaries hurting female stocks much more than male stocks (M_{male} = .07, SD = 6.89; M_{female} = -1.90, SD = 6.45; $t(56) = 2.20$, $p = .032$) (Table 4). In other words, comments about the company having low rationality or low strength were more persuasive for companies with a female name, compared to companies with a male name.

Table 4.a *The gender effect, broken down by different messages, with the whole sample*

	M	*F*	*t*	*df*	*p*
upHF[19]	5.72 *(4.77)*	5.07 *(5.46)*	*0.77*	*56*	*.445*
upHR	6.23 *(4.66)*	6.23 *(4.74)*	*0.00*	*59*	*1.000*
upLF	0.07 *(6.89)*	*-1.90 (6.45)*	*2.20*	*56*	*.032*
upLR	-1.60 *(6.62)*	-4.58 *(5.29)*	*3.14*	*59*	*.003*
downHF	2.05 *(6.59)*	3.25 *(6.13)*	*-1.32*	*59*	*.192*
downHR	2.25 *(6.50)*	1.81 *(7.15)*	*0.32*	*56*	*.751*
downLF	-5.56 *(4.87)*	-4.60 *(5.85)*	*-1.00*	*58*	*.322*
downLR	-5.52 *(4.90)*	-5.57 *(4.91)*	*0.07*	*57*	*.944*

respectively). This was driven especially by upward moving stocks that analysts had criticized as low in rationality or forcefulness ($t(50) = 2.88$, $p = .006$; $t(48) = 2.31$, $p = ..025$). For downward moving stocks, the male-female difference wasn't significant in either batch ($t(50) = 0.54$, $p = .589$; $t(48) = -1.00$, $p = .323$). Looking at overall momentum (up minus down), respondents in batch 2 expected male stocks to be higher ($t(48) = 2.17$, $p = .035$) while in batch 1 the trend was in the same direction but not significant, $t(50) = 1.31$, $p = .198$. Collapsing over both up and down moving stocks, respondents in batch 1 expected males to outperform female stocks $t(50) = 2.35$, $p = .023$; in batch 2, the trend was in the same direction but not significant, $t(48) = .77$, $p = .45$).

[19] HF: high forcefulness; LF: low forcefulness; HR: high rationality; LR: low rationality

Table 4.b *The gender effect, broken down by different messages, excluding the 18 inattentive participants*

	M	*F*	*t*	*df*	*p*
upHF	*6.00 (4.66)*	*5.76 (5.11)*	*0.26*	*45*	*.800*
upHR	6.83 *(3.80)*	6.32 *(4.60)*	*.70*	*53*	*.485*
upLF	-.09 *(7.10)*	*-2.70 (6.27)*	*2.73*	*45*	*.009*
upLR	*-2.02 (6.44)*	-4.48 *(5.34)*	*2.51*	*53*	*.015*
downHF	*2.67 (6.31)*	3.44 *(6.04)*	*-.82*	*53*	*.418*
downHR	3.00 *(6.34)*	*3.00 (6.93)*	*0.00*	*45*	*1.000*
downLF	-5.46 *(5.05)*	-4.80 *(5.74)*	*-.63*	*53*	*.529*
downLR	-5.59 *(4.93)*	*-6.67 (3.91)*	*1.71*	*45*	*.094*

Implicit and explicit endorsement of stereotypes. Explicit endorsement of stereotypes for male forcefulness was decently correlated with an individual's tendency to expect male (vs. female) stocks to be more likely to go up (collapsed over conditions), more likely to go up after having gone up in the past, and more likely to go up despite negative comments by the analyst. For rationality, the effect was only significant in the latter case and only when inattentive participants were excluded. The tables below show the correlations in these circumstances including and excluding the people who answered the attention check questions at chance levels or below.

Table 5.a *Correlation between the explicit stereotype endorsement and stock prediction, with the whole sample (n=118)*

Variable	1	2	3	4	5
1. Rationality	_				
2. Forcefulness	.37**	_			
3. Up, collapsed over conditions [a]	.04	.29**	_		
4. Up, following up	-.03	.20*	.67**	_	
5. Up, following negative comments by analyst	.12	.30**	.66**	.48**	_

[a] values are male minus female for variables 3, 4, and 5

*$p < .05$. **$p < .01$.

Table 5.b *Correlation between the explicit stereotype endorsement and stock prediction, excluding the 18 inattentive participants (n=100)*

Variable	1	2	3	4	5
1. Rationality	_				
2. Forcefulness	.39**	_			
3. Up, collapsed over conditions [a]	.13	.29**	_		
4. Up, following up	.09	.22*	.66**	_	
5. Up, following negative comments by analyst	.23*	.31**	.67**	.49**	_

[a] values are male minus female for variables 3, 4, and 5

*$p < .05$. **$p < .01$.

Implicit stereotyping. A "chauvinist" score was calculated for each participant based on their tendency to (mis)remember information in a way that favors males over females. In other words, for each of the memory questions[20], a person would earn zero points if they remembered the name-word pair correctly, one point if they mistakenly paired a female company name with a negative adjective or a male company name with a positive adjective, and negative one point if they mistakenly paired a male company name with a negative adjective or a female company name with a positive adjective. Results with the whole sample showed that the chauvinist score was not significantly different from zero (M = .038, SD = .290, $t(117) = 1.43$, p = .156) and did not positively correlate with participants' expectation for male companies to outperform female companies, collapsing across past direction ($r(118) = -0.01$, p = .931). Excluding participants who answered attention checks at the chance level (or less), the data showed a marginally significant result (M = .053, SD = .294, $t(99) = 1.79$, p = .077), possibly indicating a slight tendency to misremember male companies as having received more favorable analyst comments than female companies. However, the correlation between memory score and predictions of male outperformance was still non-significant ($r(100) = -.02$, p = .83)

2.2.3 Discussion

The following seem to be the safest conclusions:

1) Overall, respondents expect male stocks to show more momentum than female stocks. This holds regardless of whether we exclude or include participants who answered the quality

[20] The two batches of participants saw different numbers of memory questions. Batch 1 answered two memory questions while batch 2 answered the same two, plus two new memory questions, for a total of four. This was because these questions were exploratory and because in batch 1 no clear trend emerged. We decided to increase the number of memory questions in batch 2 hoping that with the increased opportunities to make mistakes, participants would be more likely to show their implicit bias if they had any. To look at the two batches together, an average (per question) 'chauvinist' score was calculated to indicate implicit stereotyping, such that those who answered 2 questions had their total score divided by 2, while those who answered 4 questions had their total score divided by 4.

control questions at chance levels or less. This male-female difference in overall expected momentum is also consistent with the archival results of Study 1.

2) These momentum effects are driven by the upward momentum effects of male stocks that had been rising vs. female stocks that had been rising. Again, this is consistent across samples and subsamples in this study, and it is also consistent with the archival results of Study 1.

3) When it comes to stocks that have been falling, respondents were not significantly more likely to predict more downward male (vs. female) movement. This was true across samples and subsamples in this study. It was also consistent with results from Study 1 when movement for falling stocks was examined.

4) There was inconsistent evidence in the present study regarding respondents' beliefs that male stocks would outperform female stocks, collapsing over the past direction. The ambiguity here parallels the ambiguity found in study 1 about the overvaluation of male stocks.

In summary, respondents predict greater momentum for male vs. female stocks, driven by the tendency for male stocks that have risen to continue going upward. Such results were found in the archival analysis of study 1 as well as the predictions of participants in the present experiment. Both forcefulness and rationality dimensions seemed to play a role as experts' comments about passivity or about emotional instability were particularly likely to "stick" when talking about female companies (vs. male companies).

Although Study 2 was conducted in a controlled lab setting with made-up stock charts and commentaries, money was at stake in the form of earning or losing bonuses (20% of the base compensation for each question); hence, subjects likely had some motivation to be correct – even if the amount of money on the line was not large in absolute terms.

2.3 General Discussion

Two studies examined the effect of grammatical gender in a monetary stakes situation: the stock market investment scenario. Study 1 analyzed real-world stock market data and showed that in Spanish-speaking countries, stocks with grammatically male names were more likely to have a larger momentum than stocks with grammatically female names, and this was driven by the tendency to rise more following a previous rising trend. When the past trend was falling, the relationship was less clear. No downward momentum was found with this sample based on how we defined momentum in this study. Instead, stocks that went down in the past still on average went back up, and there was no statistically significant difference between male and female stocks in their movement following a downward trend. Study 2 designed a stripped-down version of the investment decision-making scenario in an experimental situation while adding stock commentaries with differing types of agency messages to tackle the underlying mechanism of the gender differences. In addition to conceptually replicating results from Study 1, it further showed that for participants, stocks with a grammatically male name were better at resisting the negative commentaries about their strength and rationality while the negative labels seemed to especially stick to female stocks. Both studies found mixed evidence for the tendency of male stocks to be more overvalued than female stocks as a whole.

With these two studies we explored the Sapir-Whorfian hypothesis in a situation where linguistic effects should, at first glance, not exist because the high stakes of the situation should encourage rational behavior and bias among individual traders should be exploited away by an efficient market (Fama, 1970). Our studies suggest that such linguistic biases may in fact persist in the market over sustained periods of time. The stock market may be understood by people in terms of human agency, and this personification of stocks as having agency may be gendered.

Spanish as a grammatically gendered language offered an ideal testing ground for the gendered linguistic effect. In the following two studies we take the analysis of metaphors and the market to a different culture and use the lens of a different linguistic feature.

CHAPTER 3: CHINESE CLASSIFIER STUDIES

Besides grammatical gender, many other syntactic features provide opportunities to study languages' effect on stock momentum from different angles. Classifier use is one of them.

In Mandarin, the word *stock* is commonly used with two classifiers: 只 (zhi) and 支 (zhi). Although their pronunciation is identical, they modify two different categories of things. The former is often used with small to medium-sized animals (e.g. jaguar, monkey, bird) while the latter is often used with long, inanimate objects (e.g. chapstick, pencil, spear). On the surface, both groups seem semantically distant from the concept of stock. Although few people have studied the exact origin of using either of these classifiers with stock, some speculate that the animal zhi stemmed from the regional dialect Shanghainese while the object zhi is a miscorrection of the animal zhi by non-Shanghainese speakers (Zhang & Yang, 2012C). If so, then the usage of the animal zhi or the object zhi to modify stock is less of a conscious choice based on the context and more likely a result of the Mandarin speakers' regional dialect which permeated into their writing habits in Mandarin[21].

A conscious choice or not, both usages took wings and are widely accepted today, as shown by a large number of search results on Baidu (the counterpart of Google in China) using each classifier. However, the animal zhi is used more, with roughly twice the number of entries

[21] To the extent that classifier choice is a matter of personal writing habits, one should expect there to be existing individual difference in the preference for one Zhi over the other (pilot study 1 supports this view) and this difference should have an effect on perceived stock agency (which is what we think Study 3 shows). Moreover, the existing individual different should be measured and taken into account in the experimental study that manipulated classifier choice. In the current experiment (study 4), we did not measure this individual difference. In the future it should be measured and analyzed.

compared to the object zhi classifier ("animal zhi stock": 20,350,000; "object zhi stock: 10,650,000)[22].

With the following set of studies, I explored whether using either of the classifiers to refer to a stock would cue up resemblance of the stock to other members of the classifier category, namely, the animal zhi with animals and the object zhi with objects. One major difference between these two categories was their level of agency. I hypothesized that similar to the findings in the previous two studies with the Spanish stocks, the different expected levels of agency would lead to a differing prediction of the stock's future performance.

First, I examined to what extent the choice of one zhi over the other was a result of regional differences and to what extent people accepted both choices (pilot study 1). I also looked for preliminary evidence of the perceived agency of each of the zhi families (pilot study 2). Then an archival study used real-world stock commentary articles to test whether articles using the animal zhi predicted future trends differently than articles using the object zhi (study 3). A difference would imply that classifier choice and future prediction are correlated, but it would not speak to whether the relationship is causal, or in which direction the causal relationship runs. To answer the causation question, an experiment manipulated the usage of classifiers to examine participants' predictions of future stock performance (study 4). In studies 3 & 4, I also examined the usage of agency and object metaphors as reflected in verb choices, following Morris and colleagues' practice. Agency metaphors described the movement as a volitional action of an agent, while object metaphors described the movement as that of an inanimate object. Although active verbs did not always perfectly map onto personification or

[22] Search date: Dec.10, 2019. Both 股and 股票mean stock. The animal zhi results were the average of "只股票"and"只股"while the object zhi results was the average of "支股票" and "支股".

animacy (e.g., an inanimate object could sometimes be said to move on its own too as in 'The ball smashed the window'), it sufficed to say that animals were much more likely to move voluntarily than objects. I examined whether the verb usage provided an alternative window into the perceived agency of the stock (mediator), and whether the congruence of verb choice and classifier choice was important (moderator).

3.1 Pilot Study 1

3.1.1 Introduction

Before setting up the experiment, I first needed to establish that there was indeed some flexibility in terms of which zhi to use for stocks. Manipulating the zhi choice in the following experiments would only make sense if both usages were accepted and did not sound out of place for the participants. The Baidu search result showed some level of interchangeability for the two. However, a regional difference could still exist, such that for each regional dialect speaker the usage was quite rigid but on a country level the diversity of usage was misread as interchangeability. Moreover, if a large regional difference existed, then in the following experiments, the participants' place of origin needed to be treated as a covariate. Pilot study 1 was conducted to look for any existing regional differences in Zhi use, as well as to assess the level of interchangeability between the two classifiers when modifying stocks.

There were not many peer-reviewed sources on how the usage of the Zhi classifier to modify stocks originated and developed. One group of researchers speculated that in some southern dialects (including Shanghainese) the animal Zhi classifier was used as a general classifier suitable for many different types of nouns. Thus, when the concept of the modern stock was introduced and a classifier was needed, the animal Zhi became the default option. In particular, Shanghainese was most likely where the trend started because Shanghai housed the

largest and most influential stock exchanges in China. The Shanghai Stock Exchange was also one of the two oldest stock exchanges in the country. Consequently, Shanghai citizens had the first encounter and their way of referring to stocks influenced the rest of the country. Meanwhile, in most Northern dialects, the animal Zhi classifier had a much more restricted usage so using it for stocks didn't make much semantic sense for the Northerners. As a result, some mistook the animal zhi classifier for the object zhi classifier which shared the same pronunciation (Zhang & Yang, 2012).

Based on the limited existing information on the origin of the two usages, the most logical regional difference, should any exist, would be either between the north and the south or based on the dialect's linguistic proximity to Shanghainese. In Pilot Study 1, I divided participants' place of origin into *North* and *South* based on where the province fell in relation to the Qinling – Huaihe (meaning: the Qin Mountain and the Huai River) division line. This is a geographical division of the North and the South of China that is conventionally used in linguistic research (Szeto, Ansaldo, & Matthews, 2018). I also used geographical proximity to Shanghai as an approximation for Shanghainese's influence.

3.1.2 Method

A group of Chinese-speaking participants (n=138) from the Psychology Subject Pool and Witmart – a Chinese crowd-sourcing platform – were asked these simple questions: *"Which classifier should be used in this phrase: A. one animal zhi stock; B. one object zhi stock; C. other (please specify)"*. On the following page, they were asked "How commonly used is the phrase *one animal zhi stock*? How commonly used is the phrase *one object zhi stock*?" on a five-point Likert scale, where 1 stood for "very uncommon, very strange" and 5 stood for "very common".

They also reported which city and province they were from, and other dialects that they spoke beside Mandarin.

Distance to Shanghai. The distance to Shanghai was measured using Google Maps' *measure distance* function. For each participant, two pins were dropped on Google Maps, one in the city of Shanghai and the other in their city of origin[23]. The distance between the two cities was recorded in miles.

North vs. South. Participants' place of origin was divided into North and South based on the Qinling-Huaihe division line. Those to the north of the division line were considered northern provinces, and the rest were the southern provinces. Six provinces lie across the division line. In such cases (n=28) I used the participants' city of origin, rather than the province, to determine whether they were from the south or the north[24].

3.1.3 Results

Spearman's rho correlation coefficient was used to assess the relationship between distance to Shanghai and the choice of a classifier. There was no significant correlation between the two, r_s (130) = -0.059, p = 0.422.

A chi squared test comparing the frequency of each classifier usage for the northern and southern provinces revealed no difference (X^2 (1, 127) = 0.27, p = 0.604).

Participants overwhelmingly chose the object zhi as the more suitable classifier over the animal zhi ($n_{\text{objectzhi}}$ = 104, $n_{\text{animalzhi}}$ = 30), which was different from the popularity reflected by

[23] when the city was provided. When the city was not provided, the capital city of the province was used.
[24] The northern provinces and municipalities in this sample: Beijing, Hebei, Heilongjiang, Henan (north of Nanyang and Zhumadian), Neimenggu, Ningxia, Qinghai, Shaanxi, Shandong, Shanxi, Tianjin;
The southern provinces and municipalities in this sample: Anhui, Chongqing, Fujian, Guangdong, Guangxi, Guizhou, Henan (south of the city of Nanyang and the city of Zhumadian), Hubei (south of Shiyan and Xiangyang), Hunan, Jiangsu (south of Huai'an), Jiangxi, Shanghai, Sichuan (south of Guangyuan), Yunnan, Zhejiang. (Ma, Ren, & Xu, 2016; Zhang, Liu, Tan, and Chen, 2012)

the search engine results. Taking the absolute value of the difference between the two "how commonly used" questions gave us a measure for participants' "commitment" to one zhi over the other. A larger value stood for high commitment (i.e., thinking that one zhi was definitely right and the other was definitely odd) while a smaller value stood for higher flexibility. The value ranged between 0 (both zhi were equally common) and 4 (one zhi received a 5 while the other a 1). Sixty percent of the participants scored a 3 or lower.

3.1.4 Discussion

Pilot study 1 showed that the classifier choice for stocks was not a function of the person's place of origin. Participants' preference for the Zhi classifiers did not differ by whether they came from a northern or a southern province, nor did it differ based on the distance to Shanghai.

There were some curious findings in Pilot 1 worth noting. First, the survey showed a different picture for the popularity of the Zhi classifiers compared to the search engine results. While the search engine returned much more animal Zhi stock pages, participants in Pilot 1 overwhelmingly chose the object Zhi. One probable explanation is the effect of "now that you asked…": asking participants to explicitly decide on which Zhi classifier to use could have made them go with the more grammatically "correct" choice based on their knowledge rather than the one that they see more often, especially if the answer to the latter is ambiguous. To laypeople who are not necessarily stock traders and do not follow financial news very closely, the occasions to see stocks discussed might not be abundant, and when they do see stocks discussed and classifiers used, the focus might not be on the classifiers since they do not carry much information with them. Therefore, there might not be a very clear memory of which classifier

was used. In addition, although strictly speaking the animal Zhi classifier is the more correct usage in the sense that it appeared in one authoritative Chinese dictionary[25] as part of an example sentence for a different word, neither of the dictionary entries for the two classifiers specified their usage with stocks. When there is room for ambiguity, participants may have relied on their common sense about the most central usage for these two classifiers and quickly concluded that "animal Zhi is more often paired with animals; stocks are not animals; therefore, animal Zhi is not the correct choice".

Pilot study 2 examines the central usage for the two Zhi classifiers and demonstrates the animal Zhi's closer tie with the animal category, indicating a higher agency.

3.2 Pilot Study 2

3.2.1 Introduction

One major assumption throughout this package of studies has been that the effect of language on stock performance operated via an *agency* channel. In the Spanish studies, male stocks were seen as full of agency and consequently were more likely to continue to rise. In the current studies, the hypotheses were again that animal stocks were perceived as more agentic than object stocks, which would lead to differing expectations of future performance.

Pilot study 2 tested whether the assumption that animal zhi words were generally seen as more agentic than object zhi words held true. Two layers of the assumption were tested: 1) in a free generation task, we expected participants to list many more animals for the animal classifier, and many more objects for the object classifier. This first test may seem a no-brainer due to the

[25] The Contemporary Chinese Dictionary (6th edition) did not directly specify which Zhi classifier ought to be used with stocks. However, in the definition of a related word (Pg.1747)"being the banker" (坐庄), the animal zhi was used together with stocks to explain it: "Investors buying a large amount of one zhi (animal) stock in order to drive up the price of the stock and profit".

nomenclature of the two classifiers. However, the names "object zhi" and "animal zhi" that we picked were a simplification. In reality, the animal classifier was also used occasionally for non-animals, such as when referring to one object from a pair (e.g. one CL glove; one CL shoe). The object classifier, on the other hand, was occasionally used for more abstract concepts, some of which could be envisioned as having human-like agency (e.g. one CL band, one CL song). However, we argued that the central members of the animal classifier category were animals and the central members of the object classifier category were objects and that these were the things that more readily came to mind. This study provided data to back up this argument. 2) In a subsequent rating task, we also expected that the animal classifier phrases would be judged as more agentic than the object classifier phrases.

3.2.2 Method

There were two stages of this pilot study. In stage I, a group of Chinese-speaking participants (*n*=22) were asked to independently generate as many nouns that could be paired with either of the Zhi classifiers as possible. In stage II, the top 20 most nominated nouns for each classifier were presented to a different group of Chinese-speaking participants (*n*=31) in random order. The nouns were presented in classifier-noun phrases with their corresponding classifier (e.g. one animal Zhi chicken; one object Zhi pencil). For each phrase, participants judged the level of agency on a scale of 1 to 100. The word "agency" was translated into Chinese as "主观能动性" which is a commonly used term (but mostly in formal writing and less so in daily dialog) to refer to the ability to make one's own decisions and to act on one's own will. Moreover, a piece of definition was provided to participants in Chinese for "agency" as "having the ability to act or move voluntarily and purposefully".

3.2.3 Results

The free generation task. A total of 78 nouns were nominated as suitable to be paired with the animal Zhi, while 56 were generated for the object Zhi. Seventy-three percent of the nominated animal Zhi classifier nouns were animals. Meanwhile, none of the object Zhi classifier nouns were animals, but 7% of them were some variations of a group of people (e.g., band, army, football team), which arguably could be perceived as having agency. The rest of the object Zhi nouns were inanimate objects, making up the majority (93%).

The weighted percentage, which took the frequency of appearance of each noun into account, painted the same picture: while 78% of the occurrences of animal Zhi nouns were animal, only 6% of the object Zhi noun occurrences were about people groups.

To show the profile of the central members of each classifier family, we report the frequency of occurrence for the top 20 nouns for each classifier in Appendix C. The top five nouns for the animal classifier are all animals, while the top five nouns for the object classifier are all inanimate objects.

The rating task. A paired sample t-test comparing each participants' ratings for the animal classifier phrases with their own ratings for the object classifier phrases showed that on average, the perceived agency for the animal classifier phrases was higher than the object classifier phrases (Animal: $M = 66.18$, $SD = 21.45$; Object: $M = 34.99$, $SD = 17.91$; $t(30) = 7.49$, $p<.001$).

3.2.4 Discussion

Pilot study 2 demonstrated that the central members for the animal Zhi and the object Zhi classifiers were, as their names indicated, animals and objects, respectively. In addition, the nouns in the animal Zhi group were seen as having more agency than those in the object Zhi group. Hence, the next question in line became whether the classifier choice for a stock may be associated with the perception of that stock. If so, then would the animal classifier make the stock seem more agentic and consequently lead to more of a predicted continuation of the current trend? Or, would agency simply be interpreted as the will and the ability to "move up"? In the next study, we examined the relationship between the classifier, stocks' historical trends, and their predicted future trends using real-world stock market commentaries in Mandarin Chinese.

3.3 Study 3: an Archival study

3.3.1 Introduction

Stock commentaries provided a good opportunity to study the classifier effect on stock momentum in the real world. With pilot study 1, we have established that there was variation in classifier choice for stock and that this variation was not due to participants' places of origin. Also, close to 2/3 of the participants did not find their non-default choice unreasonable. Pilot study 2 showed that the animal classifier's associated nouns were viewed as more agentic than those of the object classifier. The key test in the present study was whether stocks used with animal classifiers were predicted to perform differently than those used with object classifiers.

Going into this study, we hypothesized that there were two possible ways in which predicted performance could be affected, and past studies have revealed support for both. A larger perceived agency could result in more attribution of the movement to enduring internal

properties rather than the external environment, and consequently, people may expect the movement to be more consistent. The Spanish grammatical gender studies found a grammatical gender effect in the upward momentum. Morris and colleagues (2007, study 1) found that an agent metaphor caused participants to expect trend continuance (i.e. larger momentum) for both upward and downward moving stocks. If animal stocks lead to predictions of trend continuity more so than object stocks, then there should be a classifier X past performance interaction in the shape of the following:

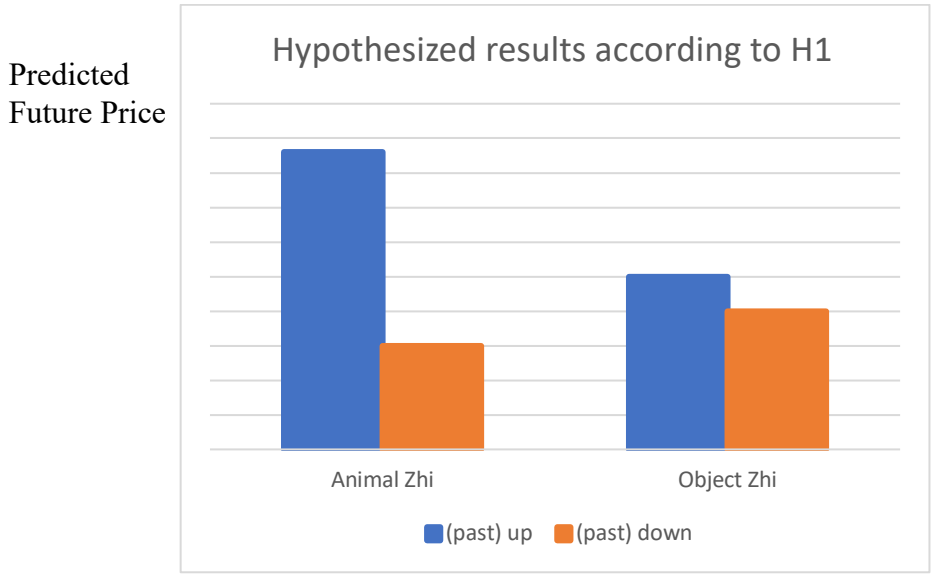


Figure 6. *Hypothesized results according to H1*

Alternatively, other studies have shown higher agency to be associated with a general tendency of 'wanting' to go up, both following previous upward trends and downward trends. For the stocks that had poor performance in the past (downward trends), one that was like an 'animal' could 'change its mind' and start going upward, while one that was like a passive 'object' could not help but continue falling. Note that this prediction was consistent with H1 when the past trend was going up but contradicts H1 when the past trend was going down. The two competing hypotheses suggest that there was a clear prediction for upward moving high

agency stocks but that prediction was less clear for the downward moving stocks. The ambiguity in downward moving stocks' predictions was consistent with the ambiguity from the results of Studies 1 and 2 with Spanish grammatical gender. Furthermore, Morris and colleagues had analyzed stock commentaries and found that agency metaphors were used more often in up-days than down-days, especially on up-days with steadier trends (2005, studies 2-4), suggesting a clear association between agency and going up. Therefore, H2 predicted that animal Zhi classifier stocks would be more likely to be predicted to go up than object zhi classifier stocks, regardless of the past movement. See Figure 7.

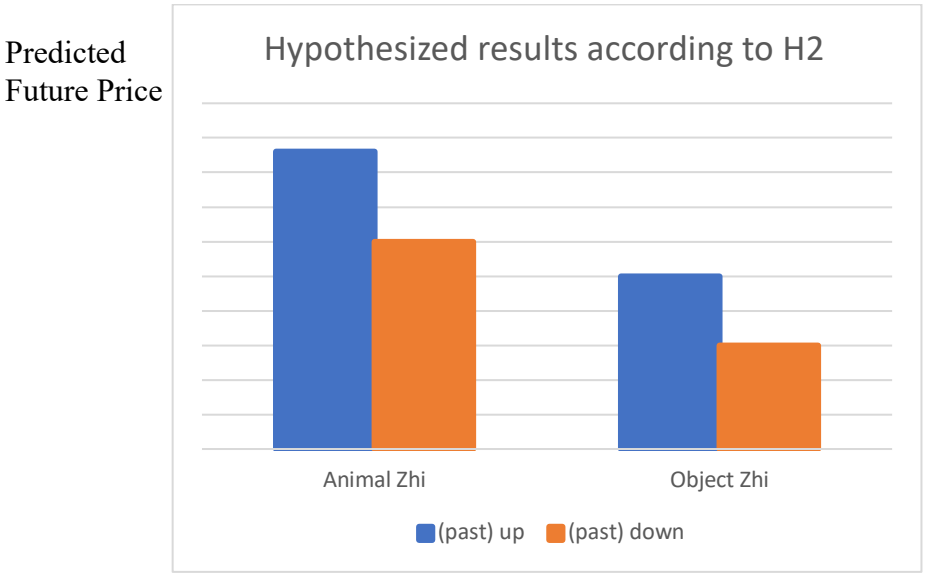


Figure 7: *Hypothesized results according to H2*

With the current study setup, we were unable to test the direction of the causal link if a link existed. Theoretically, classifiers would be the cause, such that using an animal classifier would lead people to predict stock movement in a certain way. The opposite causal direction would be less likely because classifier usage is likely a result of lasting personal writing habits unlikely to change with the context. Experiments would be needed to sort out the causal direction.

Finally, to the extent that H1 and H2 hold, they may be partially mediated by verb use (H_M). Here, active and passive verb use could provide a window into the underlying mechanism of the classifier effect. If there was indeed partial mediation by verb use, it would support the idea that differences in direction were at least partially driven by differences in the perceived agency.

3.3.2 Methods

We gathered stock market commentary articles published on the WeChat Official Account platform. WeChat is an instant messaging and social networking app developed by Tencent. It has more than 1 billion monthly active users and wide coverage of different demographics (Tencent, 2019). The WeChat Official Account is a unique type of WeChat account where individuals and businesses can push articles and updates to the WeChat users that follow them. Eighty percent of WeChat users follow at least one WeChat Official Account. Many Official Accounts gear towards sharing investment information and discussing investment strategies, including stock market commentaries. Readers gain access to the articles of different Official Accounts by following them and opting to receive their pushes. Therefore, conventional search engines such as Baidu and Google do not have access to these articles. Instead, we searched on weixin.sogou.com, a search engine specifically developed by Tencent for searching within the WeChat Official Accounts platform.

We used ‘Zhi (animal) stock” and “Zhi (object) stock” as our search terms. The former yielded around 12000 articles while the latter yielded around 7000. The search results were sorted based on the search engine’s default algorithm which took the articles’ relevance, quality (popularity), and timeliness into consideration (Su, 2017). We carried out a search weekly and

retrieved the first 30 non-duplicated articles that fitted our criteria each time. One month into the data collection, most high-ranked articles had already been collected, so to avoid getting duplicating articles we applied a date filter to limit results to the recent articles from the past week. Articles were selected if they mentioned the movement or performance of a certain stock or index as well as a prediction of future performance of the same stock or index. For coding efficiency and to have each author equally represented, when a long article discussed multiple stocks, we coded only the primary one. If all stocks were equally emphasized, we coded only the first one. We also included only one article per author. Accounts were flagged as "possible bot accounts" if we spotted multiple articles mechanically piecing together data with very similar language and paragraph structure. Such accounts were avoided. Resorting to the WeChat official accounts as the source of articles was also largely an attempt to avoid bot-written articles. Compared to articles posted on web portals that are accessible through regular search engines whose traffic relies on keyword searches and one-time visitors, the WeChat official accounts' readership, and hence profitability, rely on long-term subscriptions. Therefore, the account owners for the latter should care more about article quality rather than sheer quantity and should be less incentivized to use bots.

For each selected article, we coded for three main types of information. The first class consisted of basic information about the article, including the title, the author, the date of publication, and the classifier choice (animal Zhi or object Zhi). The second class was the price movement information, including the past movement and the prediction of future movement. These items were coded on a five-point Likert scale (1=down, 5=up). Values greater than 3 here indicate the appreciation (or expected appreciation) of price over time. The third and final class was verb/metaphor usage. We defined three groups of verb/metaphors: active verb/agent

metaphor, passive verb/object metaphor, and neutral verb/no metaphor. When describing stock movement, the metaphor could mostly be implied by the verb choice. For example, 'bounced back' implied an object metaphor, while 'leaped up' implied an agent metaphor. For the rest of this section, I will treat the concepts of verb choice and metaphor choice as equivalent. The no metaphor group consisted of descriptions using the most generic verbs (e.g. 涨 = rise; 跌 = fall) or predictions implied by making direct suggestions (e.g., 'Now is the time to buy'). One could argue that the literal meaning of the Chinese character 涨(rise) refers to the rising of water level, hence an object metaphor. Similarly, the Chinese character 跌 (fall) typically describes an agent falling rather than an object, hence an agent metaphor. However, these two verbs are used so frequently in the stock market context to describe the general rising and falling trends of the stock chart that their implications in the stock market context have likely been disassociated from their original, literal meanings outside of the stock market context. It was decided a priori that these two belonged to the neutral group. For more examples of verbs in these three categories, please see Appendix D.

Verbs for the past movement and verbs for the future movement were rated separately. For the past movement we rated the likelihood of using active, passive, and neutral verbs (E.g. 'how often does the author use active verb/agency metaphor?') A proportion estimate was entered by each coder and the entry for all three types of verbs summed up to 100. In addition, we also coded the corresponding directions of each verb type in an article (five-point Likert scale, 1=down, 5=up). We did the same for predicted future movement. Two native Chinese speakers independently coded the 331 articles. The two coders and I met weekly to discuss and resolve discrepancies.

3.3.3 Results

The coders had a perfect agreement on the binary items. Their agreement for the Likert scale items was high (for all items, ICC > 0.7)[26]. We used the average between the two coders as the final rating.

The most straightforward way of testing H1 (i.e., the animal zhi classifier use was associated with past trend continuance) vs. H2 (i.e., the animal zhi classifier use was associated with going up regardless of past trend) was to see whether there was an interaction between past direction (up or down) and classifier using a two-way factorial ANOVA. We first dichotomized the past movement direction into a two-level (down vs. up) categorical variable called *past direction*[27]. A two-way ANOVA model with classifier (animal vs. object) and past direction (down vs. up) as the two factors and the predicted direction (Likert scale, 1 = down, 5 = up) as the dependent variable showed a significant interaction between past direction and classifier ($F(1,302) = 5.14$, $p = .024$). However, the simple main effect analysis painted a different picture from H1. It showed that when the past direction was down, the animal Zhi classifier articles had significantly more positive predictions for future direction ($M = 3.54$, $SD = 0.98$) than the object Zhi classifier articles ($M = 2.84$, $SD = 1.21$, $p < .001$); however, when the past direction was up, there was no difference between classifiers ($p = .242$). This doesn't support the agency = trend continuance hypothesis in H1. See Figure 8. The data were consistent with H2 – the hypothesis that agentic types are more likely to go up – as there was an overall main effect of animal classifier stocks being predicted to rise more ($M = 3.84$, $SD = .98$) than object classifier stocks ($M = 3.57$, $SD = 1.10$, $t(329) = 2.34$, $p = .020$).

[26] ICC estimates were calculated based on a mean rating (k=2), absolute agreement, and a two-way random model, assuming both articles and coders are a sample from a larger population.
[27] n=306. 25 data points were lost due to having neutral past movement direction (i.e. past movement = 3)

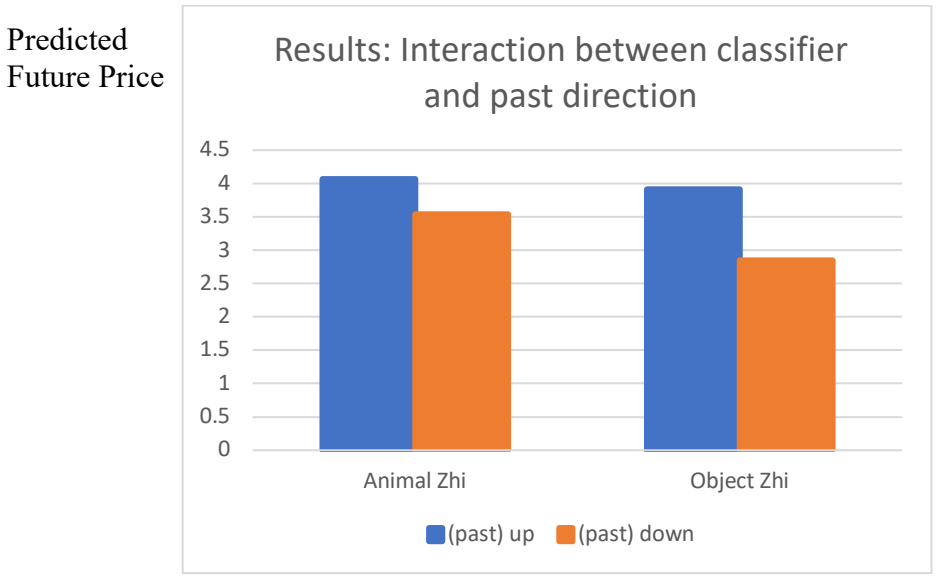


Figure 8: *Interaction between classifier and past direction*

Levene's test showed that variances for the predicted direction between different groups were not equal, $F(3, 302) = 5.388$, $p = .001$. This mattered because the sample size was also unequal (there were more articles on (past) up than on (past) down). Therefore, two approaches were taken to verify results using statistical tests without the equal variance assumption. Approach one performed tests that are more robust to the heteroscedasticity issue and found the same results[28]. Approach two tested a linear regression model with the continuous past movement variable (Likert scale, 1 = down, 5 = up) standardized and entered as a moderator[29]. As in the two-way ANOVA analysis, the predictor variable for the analysis was classifier and the outcome variable for the analysis was predicted direction. Again, the interaction between classifier and standardized past movement was found to be statistically significant ($b = -0.250$,

28 To use the Welch and Brown-Forsythe tests for equality of the means among the four groups, we flattened the factorial ANOVA down to a one-way ANOVA. Both tests were significant ($p < .001$). A posthoc Games-Howell test showed a significant difference between classifiers at past down ($p = .010$), but not at past up ($p = .574$). This confirms the results from the simple main effect F test above. Interaction was examined with the contrasts (up_animal = -1, up_object = 1, down_animal = 1, down_animal = -1). Results showed a significant interaction ($t(168) = 2.13$, $p = .035$).

29 In order to apply the Davidson-Mackinnon standard error estimator which does not assume homoscedasticity (P.22, http://claudiaflowers.net/rsch8140/Hayesprocess.pdf), the PROCESS package was installed on SPSS and used.

95% C.I. (-0.479, -0.021), p = .033). The conditional effect of classifiers on the predicted direction showed corresponding results. At low levels of past movement (Z_past_movement = -1.112; original score ≈ 2), the (conditional effect = 0.481, 95% C.I. = (0.126, 0.836), p = .008). At medium levels of past performance (Z_past_movement = 0.019; original score ≈ 3.5), the (conditional effect = 0.198, 95% C.I. = (-0.009, 0.405), p = .060). At high levels of past performance (Z_past_movement = 1.150; original score ≈ 5), the (conditional effect = -0.085, 95% C.I. = (-0.391, 0.222), p = .582). Testing the simple effect of past direction on future direction for the two classifiers separately, results show that for both animal and object classifiers, the past direction is positively correlated with future direction (Animal: $r(167) = 0.28$, $p < .001$; Object: $r(163) = 0.48$, $p < .001$), and the correlation for object classifiers is significantly stronger than that for the animal classifiers (Fisher's $z = -2.12$, $p = .034$). These results confirm the results of the two-way ANOVA analysis that there was a main effect of classifier type and a classifier X past direction interaction, such that animal stocks (vs. object stocks) were expected to go up with this effect driven by animal stocks that had been going down now being predicted to reverse direction and go up.

To test whether the verb usage in the article served as a mediator (H_M), we tested moderation mediation using PROCESS. The likelihoods of active, passive, and neutral verb usage were entered separately as the mediator. No mediation effect was found. Moreover, simple t-tests showed that the animal Zhi classifier was not associated with a larger likelihood of active verb use either in describing the past direction ($t(329) = 1.18$, p = .240) or predicting the future direction ($t(329) = 0.45$, p = .655). Nor was the object Zhi classifier associated with larger likelihood of passive verb use (past: $t(329) = -1.184$, p = .237; prediction: $t(329) = -0.069$, p = .945).

To see if the findings of Morris and colleagues would hold for the Chinese stock sample, we also explored some other characteristics of the verb choices and their relationship with stock movement direction, both past and future. Consistent with the findings of Morris et al, passive verbs were more likely to be used both to describe the past downward (vs. upward) movement ($r(330) = -.12$, $p = .035$) and the future predicted downward movement ($r(330) = -.12$, $p = .024$). On the flip side, active verbs tended (though not significantly) to be used to describe past upward (vs. downward) movement ($r(330) = .089$, $p = .107$), but not in describing the future predicted upward movement ($r(330) = -0.06$, $p = .281$).

3.3.4 Discussion

The archival study examined the usage of animal and object Zhi classifiers in real-world stock commentaries and found the two to be differentially associated with price predictions. The evidence was not consistent with H1 – that upward-moving animal stocks would continue to move up while downward-moving animal stocks would continue to move down. The evidence was consistent with H2 – that animal stocks would be predicted to move up as a main effect. The evidence was also consistent with a third perspective, however (H3). This interactive perspective reflected the idea that animal stocks that had been moving downward would reverse directions and start moving upward. This third perspective is worth some discussion.

Conceptions of agency. In the introduction to Studies 1 and 2 as well as to this third study, we have discussed agency as if it were a unitary construct. However, it may not be. Stereotypes hold that men are more agentic than women. Most people also believe that humans are more agentic than, say, rocks. However, the way men are stereotyped as more agentic than women is not the way that people are more agentic than rocks.

When we discuss differences in stereotypes of male and female agency, we are primarily thinking about agency as similar to "efficacy": the ability to achieve one's goals through acting on the world with strength, power, and means-end thinking (a focus on the goal and the rational pursuit of it, without being stifled by concerns with other goals, such as "communion." Agency vs. communion is a common way of describing male-female stereotypes (Hsu et al., 2021)). When we discuss people and objects, the former have "will": they have goals and are capable of originating action, without needing to have forces acting upon them. They are authors of their own behavior. Having goals would be consistent with animal stocks having a drive to go upward (and is consistent with a main effect that animal stocks should be predicted to rise more than object stocks); being able to act independently and without outside forces is consistent with a different type of prediction. Like participants watching the movement of simple geometric shapes (Scholl & Tremoulet, 2000; Tremoulet & Feldman, 2000; Szego & Rutherford, 2008), we perceive things as having animacy when they move on their own and in ways that run counter to physical laws: they do not follow the properties of inertia; they stop and start; they reverse direction, and so on.

In this way, agency as "independent action" may generate predictions opposite to that of predictions generated from agency as "efficacy." If independent will is manifested when objects disobey the laws of inertia, this implies that stocks with such capabilities for independent action should show changes in direction rather than simple momentum[30].

The results of Study 3 suggest that there may be something to the agency as "independent action" hypothesis. As mentioned above, the data are partially consistent with the animal stocks

[30] We should note that this prediction suggests a general tendency. There is more than one way to construe these concepts. Thus inanimate objects are subject to the law of gravity, such that objects that go up must come down. Further, animate beings with goals may resist outside forces and work to go up, despite outside forces (such as gravity) acting to push them back down.

go up hypothesis (H2) as well as consistent with the "independent action" (H3) hypothesis. Further research will be needed to see whether such hypotheses hold in out-of-sample tests and which tendency (if either) is stronger. If H3 (the "independent action" hypothesis) is supported further, it has the additional quality of being consistent with – and a possible *partial* explanation for – other phenomena that are relatively unique to the Chinese market.

As described in the introduction, momentum effects are ubiquitous in stock markets (E.g. see Park & Kim, 2014; Rouwenhorst, 1998, 1999) – except in Japan and China (Chui, Titman, & Wei, 2000; Yang, Gebka, & Hudson, 2019). In a recent review of momentum in the Chinese stock market, Yang, Gebka, and Hudson (2019) go through a number of studies as well as do their own analyses for holding and observation periods of various lengths. They report that "different studies report contradicting results, and the overall picture is that of a rather weak price reversal rather than momentum effects" (p. 98). In their own analysis, they find evidence for reversals as well, though results depend somewhat on the time the study is conducted as well as the length of the holding period.

A hypothesis of reversals driven by "independent action" would also be consistent with findings by Ji, Zhang, & Guo (2008) and the work of Rieger et. al (2022) showing that Chinese (but not the control groups from other countries) participants were more likely to expect stocks to show reversals or to show no effect of past trends. Previous work on Asian dialecticism and predictions of reversal have relied heavily on scenarios that involve some combination of human agency and perhaps luck (predictions about income, happiness, relationships, performance, and so on) (Ji, Nisbett, and Su, 2001; Ji, 2008). Such findings are consonant with why animate objects might show reversals, but inanimate objects would show momentum (inertia), all other things equal. Presumably, Chinese respondents do not reject Newton's first law – the law of

inertia -- that objects at rest tend to stay at rest while objects in motion continue in that motion unless acted upon by some force[31]. "Will," in the sense of independent action by animate beings, can counter inertia.

3.4 Study 4: an Experiment

3.4.1 Introduction

Study 3 found some evidence of the classifier effect on the expected future price movement in the real world. Because the classifier usages were measured rather than manipulated, it was not suited to tackle the causal direction question. One direction is more likely than the other. Natural variation exists in terms of which classifier a person habitually uses or considers to be the correct usage based on their knowledge, and this habit is less likely to change based on momentary context. Therefore, it is more likely that the classifier is the cause, and the perceived stock movement direction is the result, rather than the other way around. Study 4 used an experimental design to test this causal link.

Based on findings from Study 3, we hypothesized an interaction between the past movement direction (up vs. down) and classifier, such that while object stocks would be more likely to show momentum (i.e. following 'inertia'), animal stocks would go up regardless of past direction (i.e. following 'will'). In addition, in order to highlight the use of animal vs. object classifiers in sentences, we ran another study concurrently where primes of animals or objects were introduced at the beginning of the experiment to serve as a 'nudge'. Two types of animals (high vs. low force. E.g., jaguar vs. squirrel) and two types of objects (high vs. low force. E.g.,

[31] Again, context is important: gravity may be a force acting on animate and inanimate objects, though only animate objects can independently initiate action to counter that force.

arrow vs. pencil) were primed but this distinction between high and low force types had little effect and so this factor ended up being collapsed over. For ease of comparing prime effects, we combined the prime and no prime studies below and present the results in a 2 (stock classifier: animal zhi vs. object zhi) * 2 (previous direction of stocks: up vs. down) * 3 (prime: no prime vs. object prime vs. animal prime) format.

3.4.2 Methods

Participants. Only native and fluent Mandarin Chinese speakers would be able to distinguish the subtlety of classifier choice and the potential implied mental imageries of different agency levels. Therefore, we recruited Chinese speakers and required that they either be native speakers (speaking Mandarin Chinese as their first language), or non-native but completely fluent in speaking, listening, reading, and writing. Because of the scarcity of Mandarin Chinese speakers in the Psychology subject pool and Amazon Mechanical Turk, we also used Prolific, Witmart, and Wechat discussion groups as alternative sources of participants. We gathered data from a total of 630 participants, of which 252 were male and 305 were female. The rest declined to indicate their gender.

Procedure. We ran two studies concurrently, one with prime and one without prime.

The priming task varied slightly based on the platform. With bilingual participants (SONA, Mturk, and Prolific), priming was achieved via a translation task and dressed up as an assessment of their Chinese fluency. With monolingual participants from Witmart and Wechat discussion groups, a *scrambled sentence* task with the same sentences as the translation was used and the cover story was that this was an assessment of cognitive ability. All participants saw seven sentences, five of which incorporated the classifier at least twice, the other two were filler

items. To encourage accuracy and increase the power of the priming, the corresponding classifier appeared more than once in each sentence, and the first appearance of the classifier at the beginning of the sentence was provided for participants. Participants only needed to follow the usage of the given classifier. The animals and objects being primed also varied on levels of force, resulting in four different priming conditions: high force animal (e.g., "*These three wolves are standing by the bush far from those two wolves*"), low force animal ("*These three squirrels…*"), high force object (*"He bought two arrows and put one in the box*"), and low force object (*"He bought two pencils and…"*). The rest of the study was identical between the prime and the no-prime version.

Participants read about eight different Chinese stocks. The information available for each company included: a short and neutral news excerpt; a one-sentence description of the stock's performance last year; a price chart of the stock that matched the description. These pieces of information were allegedly taken from Sohu Finance - a reputable Chinese financial news source similar to Yahoo Finance - when in reality they were created by the researchers. The ambiguous news was not manipulated; it was added to create verisimilitude and did not involve comments by analysts. For each stock, the participants were asked whether they would buy or sell the stock, and how certain they were about their decision.

Within person across stocks, the verbs used in the one-sentence description of the stock performance last year also varied in terms of agency, but since this also had no effect on the results, it was collapsed over in the results section for simplicity of presentation. (There were eight different verbs, one for each of the eight stocks, of which two were up action (leap up 跃居高位, march up 走高), two were down action (dive 跳水，retreat from defeat 节节败退), two

were up passive (bubble up 快速上浮，was hoisted up 被拉升), and two were down passive (was flushed down 一泻千里，was caught in the downdraft 陷入下跌的漩涡).

The pairing of the neutral news excerpt and the trend description was counterbalanced. In the body of the news excerpt as well as the trend description, whenever the stock was mentioned, it was consistently paired with one of the classifiers: the animal Zhi or the object Zhi, hence the classifier usage was a between-person variable. This setup is different from the one involving grammatical gender in Study 2 where the gender of company names varied within-person, because classifier and grammatical gender are functionally different in the languages. Grammatical gender is an inherent property of the company's name, so it was sensible to vary the gender within the eight companies that were presented to a subject. However, the usage of the animal or object zhi classifier is a matter of habit and it makes little sense to be switching back and forth between animal and object zhi classifiers within-person.

We also collected information on participants' demographics including their gender, age, level of education, experience with the stock market, financial literacy, where in China they come from (when applicable), and whether they spoke any other languages or regional dialects besides Mandarin Chinese.

It is vital that participants read the stimuli carefully and make decisions with consideration. Therefore, we adopted a compensation scheme similar to what's used in Study 2 to incentivize 'good performance' in the stock trading section. Subject pool participants received course credits and were incentivized to think seriously and perform well in the prediction by a raffle draw. They were told that the more they predicted correctly (i.e., bought and the stock later rose, or sold and the stock later fell), the more raffle tickets they would win to be entered into a draw at the end of the data collection phase to win $20. They were later debriefed that there is no

correct answer and that everyone will have an equal chance in the raffle draw. Participants from the other platforms received $0.5 (or equivalent Chinese Yuan) as compensation and were incentivized to perform well in the stock predictions because we told them that each correct answer would earn them $0.1 while each wrong answer would lose them $0.1. In the end, they were debriefed, and each participant received $0.2 as a bonus.

In summary, this study employed a 2 (animal vs. object classifier for target) X 3 (animal prime, object prime, no prime) X 2 (up-down) X 2 (active vs. passive verbs) X 2 (low agency prime vs. high agency prime) factorial design. The first two factors are between-person factors while the rest are within. We ended up collapsing over the latter two factors because they did not make a difference.

3.4.3 Results

Data from the prime and the no prime studies were merged together to show the effect of prime. A composite variable was created by multiplying participants' dichotomous response (sell = -1, buy = 1) with the zero-centered 10-point Likert scale item on their confidence (completely guessing = -4.5, very certain = 4.5) to serve as the dependent variable: the future stock movement prediction. Positive value indicates expecting price to appreciate / move upwards.

Because direction was a within-subject variable while classifier and prime were between-subject variables, a mixed ANOVA testing the three-way interaction between past direction (up vs. down), prime (no prime vs. object prime vs. animal prime), and classifier (object classifier vs. animal classifier) was conducted to examine their effects on future stock movement prediction. The omnibus F for the overall 3-way interaction was $F(2, 625) = 2.37$, $p = .094$. Unlike in the

previous studies, this experiment provided no evidence for the main effect of animal classifier stocks being more likely to go up ($F(1, 625) = 0.348$, $p = .56$).

For ease of interpretation (and for consistency with Study 2), we conceptualized momentum overall as the difference between up and down – *up-down* (read: 'up minus down') – where a larger positive difference indicates larger momentum. There was a significant momentum effect for all conditions except the animal classifier when primed with either animal or object sentences (Table 6).

Table 6. *Momentum effect (up – down) for the different prime conditions with animal and object classifiers*

Condition	*M*	*SD*	*N*	95% CI	*p*
Animal Classifier					
1. Object prime	.07	2.13	68	[-0.43, 0.58]	.78
2. No prime	.75	2.08	168	[0.44, 1.07]	<.001
3. Animal prime	.06	2.39	77	[-0.47, 0.59]	.83
Object Classifier					
4. Object prime	.49	1.75	77	[0.10, 0.88]	.02
5. No prime	.49	2.41	173	[0.13, 0.85]	.008
6. Animal prime	.63	2.14	68	[0.12, 1.14]	.02

Furthermore, as shown in Table 7 (below), the lack of momentum for animal classifier stocks with primes was driven primarily by a reversal of downward movement (following previous downward trends) and less so by a lack of upward momentum (following previous upward trends): following previous downward trends, there was a significant contrast between

animal classifier stocks with a prime and the rest of the conditions (conditions 1&3 vs. the rest, $t(625) = 2.49$, $p = .013$). This is consistent with the findings of Study 3 where using an animal classifier may be a reflection of – or result in – seeing stocks as having 'independent will' like an animal, implying that the stocks could resist downward inertia. Following previous upward trends, though animal classifier stocks with a prime tended to show less momentum, this difference was not significant (contrasting 1&3 vs. the rest: $t(625) = 1.45$, $p = 0.147$). This stunted upward momentum (although not significant) is unlike what we saw in Study 3 where the upward momentum was completely unaffected by the classifier. Moreover, this effect was not only found in the animal prime + animal classifier condition (simple effect of animal classifier stocks in the animal prime condition vs. no prime condition: Up-Down $t(243) = -2.20$, $p = .029$; Down $t(243) = 2.01$, $p = .04$; Up $t(243) = -1.55$, $p = .123$), but also in the object prime + animal classifier condition (object prime vs. no prime condition: Up-Down $t(234) = -2.25$, $p = .026$; Down $t(234) = 2.17$, $p = .03$; up $t(234) = -1.15$, $p = .253$). Thus, animal-classifier stocks reversed course where they should have – on stocks that had been moving downward after an animal prime. But animal-classifier stocks also reversed course where they should not have (at least according to Study 3's analysis – on stocks that had been moving upward (nonsignificantly) and after an object prime) [32].

[32] The outlier-removed (based on 1.5 IQR rule) analysis shows a similar overall picture. The three-way interaction was significant ($F(2, 583) = 3.259$, $p = .039$), the pattern of the cell means is similar to the complete data and the results remain similar. The column with the up minus down difference score has slightly different outliers than when up and down are considered separately; thus the third column is not the same as the first minus the second. When analyses are performed on the third column, the two-way interaction between prime condition and classifier was also significant ($F(2, 600) = 4.34$, $p = .013$).

	Object Classifier			Animal Classifier		
	up	down	up-down	up	down	up-down
Object Prime	.28 (.97)	-.13 (1.03)	.47 (1.47)	.30 (1.08)	.01 (1.23)	.12 (1.69)
No Prime	.16 (1.05)	-.21 (1.16)	.37 (1.83)	.26 (1.07)	-.25 (1.14)	.63 (1.78)
Animal Prime	.38 (1.22)	-.16 (1.10)	.65 (1.82)	-.08 (1.26)	.12 (1.17)	-.09 (1.93)

Table 7. *$Momentum_{up}$ and $Momentum_{down}$ for the different prime conditions with animal and object classifiers*

Condition	M_{up}	SD_{up}	M_{down}	SD_{down}	N
Animal Classifier					
1. Object prime	.19	1.35	.12	1.55	68
2. No prime	.41	1.40	-.34	1.25	168
3. Animal prime	.10	1.48	.04	1.45	77
Object Classifier					
4. Object prime	.33	1.10	-.15	1.29	77
5. No prime	.22	1.39	-.27	1.46	173
6. Animal prime	.38	1.41	-.25	1.34	68

3.4.4 Discussion

Results from Study 4 were partially consistent with the *animal = reversal at downtrend* hypotheses discussed in Study 3 (H3) in that the effect held but only when primes were used. The animal primes highlighted the animal classifier Zhi in that participants first used this classifier for sentences involving animals before considering the stocks. This may have made the 'animal-ness' of the animal Zhi classifier more salient in the minds of participants, consequently leading to the prediction of reversals (especially reversals from downtrends) rather than trend continuity. Surprisingly, object primes also had this same effect on animal classifier stocks. We did not have an a priori hypothesis that this effect would occur. To speculate, this could be due to a 'contrast' effect where being reminded of the 'object-ness' of the object Zhi classifier somehow made the animal-ness of the animal Zhi classfier pop; perhaps the juxtaposition of object Zhi and animal Zhi made the classifiers and their associated categories more salient. The

contrast hypothesis seems plausible. In English, for example, reading sentences about litter in a parking lot may make the word "litter" pop when one later hears about a litter of puppies (an example of a homograph); reading about guerilla warriors may make the word "gorilla" pop when one later hears about a band of gorillas (a homophone); and so on. This is speculation but future work can investigate this possibility.

Study 4's results differed from those of Study 3 in that the experiment showed a trend of stunted upward momentum (although not significantly) for the animal classifier stocks with primes. In comparison, the real-world data in Study 3 showed no difference in the upward momentum between the animal and the object classifiers. Also, the experiment found no sign of an overall 'animals go up' effect.

Perhaps future studies will find significant reversals of upward momentum for animal stocks. If so, stock commentaries on social media may not be the ideal place to look. Stock commentators have the incentive to write about promising stocks that will likely go up in the future because they are what are profitable and what readers care about (note: the Chinese stock market does not allow shorting). Future studies exploring the possibility of reversals for upward momentum will likely need larger n for experiments and an archival study would need data from people less incentivized to bullishly tout stocks.

Taken together, the present data are consistent with the other research. In the archival study, there is evidence suggestive of a classifier effect on stock market predictions (Study 3), but the experiment is also consistent with there being boundary conditions: in Study 4, subjects needed to be primed with either animal or object Zhi sentences to produce the classifier effect on predictions.

CHAPTER 4: GENERAL DISCUSSION

Four studies examined the hypothesis that arbitrary linguistic factors, such as the grammatical gender and the classifier usage, can affect the perception of stock performance and sway individuals' investment decisions.

Two studies out of the four (one archival study, one experiment) tested the effect of grammatical gender in Spanish and found significant results in favor of this hypothesis: companies with grammatically male names were expected to go up more than those with grammatically female names. This was especially true when following an upward trend, as reflected by male companies displaying a significantly larger upward momentum than female companies. The results following downward trends were murkier. The archival study showed a nonsignificant tendency for male stocks to reverse downward trends more than female stocks; the experiment did not. There was mixed evidence as well for male stocks being overvalued. In the archival study, male companies showed marginally higher valuation (P/E, BV/MC) than female companies. In the experiment, male companies were overall expected to outperform female companies, though this held only when confirmed nonreaders were eliminated. The experiment also investigated a potential underlying mechanism for the gender effect. It found evidence supporting an agency argument: first, criticisms of irrationality and weakness tended to "stick" to female companies more than they did for male companies. Second, the more one believed that men were stronger/more forceful than women, the more likely one expected male-named companies to go up.

The other two studies (one archival study, one experiment) focused on classifier usage in Mandarin Chinese and how it affects the perception of stocks in the Chinese stock market. Results showed a classifier * past direction effect. In the archival study of stock commentaries,

whereas upward momentum was expected regardless of the classifier, downward momentum was only predicted when stocks were described with an object classifier Zhi (as if following inertia). When stocks were described with an animal classifier Zhi, they were more likely predicted to reverse the current trend and go higher (as if following a will to go up). The experiment showed a similar finding – but only when subjects had first been primed with sentences that used either the animal Zhi prime or object Zhi prime. Why the object prime produced a contrast effect is not entirely clear at the moment.

Conceptions of agency, revisited. We have demonstrated that a different level of agency from some arbitrary linguistic components may be driving the differing predicted outlook for stock performances. Although agency has been discussed as a unifying theme that strings together the Spanish and the Chinese studies, the results of the Chinese studies call for a closer look at the different conceptions of agency. In the Spanish studies, agency manifested through gender stereotypes of males and females. Males are seen as more forceful and rational. The master of his own action, a willful man stereotypically has a strong and clear sense of where he is going, and as a sensible agent, wants to go up. This explains why we mainly saw effects in the upward momentum in the Spanish studies. In the Chinese study, however, the distinction was made between animals and objects. Animals also show more agency in their behavior in the sense that they can initiate actions, but agency differences between animals and objects are not like agency differences between men and women. Rather, objects lack agency on a more fundamental level, such that they are always subjected to the total influence of outside forces (inertia, for example). One prominent outside force that agentic animals can resist while objects cannot would be gravity, the direction of which points downward. Thus, we saw stocks with animal classifiers ‘resisting gravity,’ reversing their downward trend more easily while object

classifier stocks had no way to resist but to fall. An animal can run down a hill, change its mind, and run back toward the top; a rock or a snowball cannot.

These are by no means the only ways that agencies can be contrasted. To be an agent means one can 'influence intentionally one's functioning and life circumstance' (Bandura, 2006). For example, when one says 'humans are more agentic than animals' they might be emphasizing the humans' abilities to make plans and carry through vs. the impulsive or reflexive nature of animals. Adults (usually) have more agency than children. Bosses usually have more agency than subordinates. Citizens of democracies can have more agency than subjects of repressive regimes. All of these contrasts – men vs. women, animals vs. objects, humans vs. animals, adults vs. children, bosses vs. subordinates, citizens vs. subjects – point to somewhat different ways agency can be manifested. Thus, the implications of projecting greater "agency" onto one entity or another may also depend on the implicit contrast that is being made.

4.1 Limitations and Future Direction

One limitation of the present studies is the restricted source of participants in the experiments. We mainly relied on MTurk and other crowdsourcing platforms to recruit participants for our experiments. The quality of the data was compromised as a result, as we did find a number of participants going through the study without paying much attention and their responses influenced some of the results and conclusions (as in Study 2). Moreover, the percentage of participants who have had prior experience investing in the real-world stock market in our sample is not very high (41.5% for study 2 – Spanish grammatical gender; 61% for study 4 among the no-prime group; the with-prime group was not asked this question). It is reasonable to suspect that one's exposure to stock market investment and exposure to how stocks

are described in commentaries may moderate the language effect. For example, the language effect could be weaker when testing seasoned investors because they may have more information or existing hunches to base decisions upon other than what the linguistic elements prompt. Although our current data do not support this distinction[33], future studies could develop more meticulous measures of participants' past investment experience and financial literacy and assess it more systematically, or they could focus more on recruiting seasoned investors and see if effects replicate.

4.2 Conclusion

While previous research has demonstrated that language influences perception, memory, and behavior in various lab and natural settings, the present studies suggest that such an effect is also present in a high-stakes situation: predictions about stock price movement can be influenced by arbitrary linguistic factors such as the grammatical gender of the stock's name and the classifier that is used to describe the stock. We argue that characteristics implied by grammar (i.e. gender and animacy) carry over to influence people's perception of the stocks and consequently lead to different predictions of future stock performance. Further research can investigate other linguistic effects and examine whether different aspects of agency manifest themselves as people try to make sense of complex and unpredictable market mechanisms.

[33] In Study 2 (Spanish study), a composite score to gauge financial literacy (based on whether participants correctly answered three basic financial literacy questions (Lusardi & Mitchell, 2011, score ranged from 0 – 3) and a composite score to gauge experience with stock investment ("Having invested in the stock market in the past" = 1 point; "currently investing in the stock market" = 1 point. Range = 0 – 2 points) were calculated. Not surprisingly, the two were positively correlated ($r(118) = .21$, $p = .02$). But neither were correlated with any of the momentum measures or the gender difference in momentum.
In Study 4 (Chinese experiment study), the no prime group also answered questions about financial literacy and past experience in the stock market. Again, not surprisingly, financial literacy and experience were positively correlated ($r(341) = .41$, $p < .001$. With the whole sample, neither financial literacy nor experience were correlated with any of the momentum measures. When outliers were removed, financial literacy was positively correlated with upward momentum ($r(341) = .12$, $p = .033$). However, financial literacy and experience did not interact with classifier to predict any of the momentum measures.

REFERENCES

Abarbanell, J. S., & Bushee, B. J. (1997). Fundamental analysis, future earnings, and stock prices. *Journal of accounting research*, *35*(1), 1-24.

Allan, K. (1977). Classifiers. *Language*, *53*(2), 285-311.

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, *68*(3), 929-985.

Asness, C., Frazzini, A., Israel, R., & Moskowitz, T. (2014). Fact, fiction, and momentum investing. *The Journal of Portfolio Management*, *40*(5), 75-92.

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on psychological science*, *1*(2), 164-180.

Boroditsky, L. (1999). First-language thinking for second language understanding: Mandarin and English speakers' conceptions of time. *Proceedings of the 21st Annual Meeting of the Cognitive Science Society*, Vancouver, BC

Boroditsky, L., Schmidt, L., & Phillips, W. (2003). Sex, Syntax, and Semantics. In *Language in mind: Advances in the study of language and cognition*, ed. D. Gentner & S. Goldin-Meadow, pp. 61- 80. Cambridge University Press.

Central Intelligence Agency. (2015). Languages. In The world factbook. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/fields/402.html

Chen, K. (2013). The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets, *American Economic Review*, 103:2.690-731

Chui, A, Titman, S and Wei, K.C., 2000, Momentum, ownership structure, and financial crises: An analysis of Asian stock markets, working paper, University of Texas at Austin.

CNBC. (2018) Markets: Expect more market shocks in China as investors figure out what they're doing, ex-central bank chief says. Retrieved from https://www.cnbc.com/2018/09/11/chinas-stock-market-is-still-immature-so-volatility-is-to-be-expected-former-pboc-governor-says.html

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998). "A Theory of Overconfidence, Self Attribution, and Security Market Under- and Over-Reactions." Journal of Finance 53, 1839-1885.

Emirbayer, M., & Mische, A. (1998). What is agency?. *American journal of sociology*, *103*(4), 962-1023.

Frazzini, A. (2006). "The Disposition Effect and Underreaction to News." Journal of Finance 61.

Gao, M. Y., & Malt, B. C. (2009). Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes*, 24(7-8), 1124-1179.

Goyal, A., & Wahal, S. (2015). Is momentum an echo?. *Journal of Financial and Quantitative Analysis*, 50(06), 1237-1267.

Graham, B. (1965). *The intelligent investor*. Prabhat Prakashan.

Heider, E. R. (1972). Universals in color naming and memory. *Journal of experimental psychology*, 93(1), 10.

Hicks, D. L., Santacreu-Vasut, E., & Shoham, A. (2015). Does mother tongue make for women's work? Linguistics, household labor, and gender identity. *Journal of Economic Behavior & Organization*, *110*, 19-44.

Holodny, E. (2017, November 10). *Isaac Newton was a genius, but even he lost millions in the*

stock market. Business Insider. Retrieved September 13, 2021, from https://www.businessinsider.com/isaac-newton-lost-a-fortune-on-englands-hottest-stock-2016-1.

Hunt, E., & Agnoli, F. (1991). The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review*, *98*(3), 377.

Hsu, N., Badura, K. L., Newman, D. A., & Speach, M. E. P. (2021). Gender,"masculinity," and "femininity": A meta-analytic review of gender differences in agency and communion. Psychological Bulletin, 147(10), 987.

Huettig, F., Chen, J., Bowerman, M., & Majid, A. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye gaze behavior but only during linguistic processing. *Journal of Cognition and Culture*, 10(1-2), 39-58.

Imai, M., & Saalbach, H. (2010). Do Classifier Categories Influence Conceptual Structures?. Words and the Mind, 138.

Imai, M., Saalbach, H., & Stern, E. (2010). Ae Chinese and German children taxonomic, thematic, or shape biased? Influence of classifiers and cultural contexts. *Frontiers in Psychology*, 1, 194.

Imai, M. , Schalk, L., Saalbach, H., & Okada, H. (2014). All giraffes have female-specific properties: Influence of grammatical gender on deductive reasoning about sex-specific properties in German speakers. Cognitive Science, 38(1), 514–536. doi : 10.1111/cogs.12074

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of Finance,* 45(3), 881-898.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance* 48, 65–91.

Jegadeesh, N., & Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of Finance* 56, 699-720.

Ji, L. J., Nisbett, R. E., & Su, Y. (2001). Culture, change, and prediction. *Psychological science*, *12*(6), 450-456.

Ji, L. J. (2008). The leopard cannot change his spots, or can he? Culture and the development of lay theories of change. *Personality and Social Psychology Bulletin*, *34*(5), 613-622.

Ji, L. J., Zhang, Z., & Guo, T. (2008). To buy or to sell: Cultural differences in stock market decisions based on price trends. *Journal of Behavioral Decision Making*, *21*(4), 399-413.

Kann, K. (2019). Grammatical gender, neo-whorfianism, and word embeddings: A data-driven approach to linguistic relativity. arXiv preprint arXiv: 1910.09729.

Konishi, T. (1993). The semantics of grammatical gender: A cross-cultural study. *Journal of psycholinguistic research*, *22*(5), 519-534.

Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *The Quarterly Journal of Economics*, *105*(1), 1-28.

Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, 83(3), 265-294.

Lindsey, D. T., Brown, A. M., Brainard, D. H., & Apicella, C. L. (2015). Hunter-gatherer color naming provides new insight into the evolution of color terms. *Current Biology*, *25*(18), 2441-2446.

Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science, 16,* 601-608.

Lusardi, A., & Mitchell, O. S. (2011). Financial literacy around the world: an overview. *Journal of pension economics & finance*, *10*(4), 497-508.

Ma, X., Ren, X., & Xu, J. (2016). The difference of collectivism between north and south China and its cultural dynamics. *Advances in Psychological Science*, *24*(10), 1551-1555.

Menkhoff, L., Sarno, L., Schmeling, M., & Schrimpf, A. (2012). Currency momentum strategies. *Journal of Financial Economics*, *106*(3), 660-684.

Mickan, A., Schiefke, M., & Stefanowitsch, A. (2014). Key is a llave is a Schlussel: A failure to replicate an experiment from Boroditsky et al. 2003. *Yearbook of the German Cognitive Linguistics Association,* 2(1), 39-50.

Morris, M. W., Sheldon, O. J., Ames, D. R., & Young, M. J. (2007). Metaphors and the market: Consequences and preconditions of agent and object metaphors in stock market commentary. *Organizational Behavior and Human Decision Processes*, 102(2), 174-192.

Moskowitz, T. J., & Grinblatt, M. (1999). Do industries explain momentum? *The Journal of Finance*, *54*(4), 1249-1290.

Murphy, J. J. (1999). *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. Penguin.

Nicaise, L. (2014). The market as a rational and responsible human being: Measuring the impact of metaphor on financial decisions. *Metaphor and the Social World*, 4(1), 90-108.

Norman, J. (1988). *Chinese*. Cambridge University Press.

Park, K.-I., Kim, D. (2014). Sources of momentum profits in international stock markets *Account Finance.* 54, 567–589.

Phillips, W., & Boroditsky, L. (2003). Can quirks of grammar affect the way you think? Grammatical gender and object concepts. In *Proceedings of the 25th annual meeting of the Cognitive Science Society* (pp. 928-933). Mahwah, NJ: Lawrence Erlbaum Associates.

Prewitt-Freilino, J. L., Caswell, T. A., & Laakso, E. K. (2012). The gendering of language: A comparison of gender equality in countries with gendered, natural gender, and genderless languages. *Sex roles*, *66*(3), 268-281.

Rieger, M. O., Wang, M., Phan, T. C., & Gong, Y. (2022). Trend following or reversal: Does culture affect predictions and trading behavior?. *Global Finance Journal*, 100769.

Roberson, D., Davies, I., & Davidoff, J. (2002). Color categories are not universal: Replications and new evidence. *Theories, technologies, instrumentalities of color: Anthropological and historiographic perspectives.* (pp. 25-35) University Press of America, Lanham, MD.

Rouwenhorst, K. G. (1998). International momentum strategies. *The Journal of Finance*, *53*(1), 267-284.

Rouwenhorst, K.G. (1999). Local return factors and turnover in emerging stock markets. J. Finance 54, 1439-1464.

Saalbach, H., & Imai, M. (2007). Scope of linguistic influence: Does a classifier system alter object concepts?. *Journal of Experimental Psychology*: General, 136(3), 485.

Saalbach, H., Imai, M., & Schalk, L. (2012). Grammatical gender and inferences about biological properties in German-speaking children. *Cognitive science*, *36*(7), 1251-1267.

Sato, S., Ottl, A., Gabriel, U., & Gygax, P.M. (2022). Assessing the impact of gender grammaticization on thought: A psychological and psycholinguistic perspective.

Schmitt, B. H., & Zhang, S. (1998). Language structure and categorization: A study of classifiers in consumer cognition, judgment, and choice. *Journal of Consumer Research*, 25(2), 108-122.

Scholl, B.J., & Tremoulet, P.D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4, 299–309.

Sera, M. D., Berge, C. A., & del Castillo Pintado, J. (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development*, *9*(3), 261-292.

Slobin, D. (1996). From "thought and language" to "thinking for speaking." In J. Gumperz & S. Levinson (Eds.), Rethinking linguistic relativity. Cambridge, MA: Cambridge University Press, 70-96.

Su, R.X., (2017, February 6). Sogou Weixin's programmer answering why search engines do not support time sort. *Zhihu.* https://www.zhihu.com/question/55206965

Szego, P.A., & Rutherford, M.D. (2008). Dissociating the perception of speed and the perception of animacy: a functional approach. *Evolution and Human Behavior*, 29, 335–342.

Szeto, P. Y., Ansaldo, U., & Matthews, S. (2018). Typological variation across Mandarin dialects: An areal perspective with a quantitative approach. *Linguistic Typology*, *22*(2), 233-275.

Tai, J. H. (1994). Chinese classifier systems and human categorization. In honor of William S.-Y. Wang: Interdisciplinary studies on language and language change, 479-494.

Tencent. (2019). Tencent 2019 Q3 Financial Report. https://static.www.tencent.com/uploads/2019/11/13/8f86f757dd226126dd13d60664e2404b.pdf

Tremoulet, P.D., & Feldman, J. (2000). Perception of animacy from the motion of a single object. *Perception*, 29, 943–952.

Twain, M. (2007). A tramp abroad. Modern Library.

Tzeng, O. J., Chen, S., & Hung, D. L. (1991). The classifier problem in Chinese aphasia. *Brain and language*, *41*(2), 184-202.

Wittgenstein, L., & S., H. P. M. (2010). *Philosophische Untersuchungen = philosophical investigations*. Wiley-Blackwell.

Yang, Y., Gebka, B., & Hudson, R. (2019) Momentum effects in China: A review of the literature and an empirical explanation of prevailing controversies. *Research in International Business and Finance*, vol. 47(C), 78-101

Zhang, J., Liu, X., Tan, Z., & Chen, Q. (2012). Mapping of the north-south demarcation zone in China based on GIS. *Journal of Lanzhou University (Natural Sciences)*, *48*(3), 28-33.

Zhang, S., & Schmitt, B. (1998). Language-dependent classification: The mental representation of classifiers in cognition, memory, and ad evaluations. *Journal of Experimental Psychology: Applied*, 4(4), 375-385. doi:http://dx.doi.org/10.1037/1076-898X.4.4.375

Zhang, Q. R., Shu, H., & Liu, Y. Y. (2011). A Cognitive View on the Study of Individual Classifier in Chinese. *Advances in Psychological Science*, 19(4), 510-520.

张军，& 杨振华. (2012). 股票是一 “只” 还是一 “支”? [For stocks, should the animal Zhi or the object Zhi be used?]. *语文建设*, (1), 50-51.

APPENDIX A. EXAMPLES OF COMPANY NAMES IN STUDY 1

Male	Female	Ambiguous
Suelo Petrol	Empresa de Generacion Electric	West Indies Rum Distillery Ltd
Banco del Caribe	Corp de Gestion Minera SA	Colina Holdings Bahamas Ltd
Corp Grupo Quimico CA	Inmobiliaria IDE SA	TGLT SA
Productos EFE SA	Empresa Agricola La Union SA	Sacyr
Montevideo Refrescos SA	Santa Luisa	Prim
Supermercados Peruanos SA	Sociedad Minera Corona SA	Imaginarium
Banregio Grupo Financiero SAB	Financiera TFC SA	Let's GOWEX
El Puerto de Liverpool SAB	Propafisa SAECA	Applus Services
Vocento	Farmacias Benavides SAB de CV	AB-Biotics

APPENDIX B. STUDY 2 NEWS, COMMENTS, AND GRAPHS

News[34]

1. La Nibsa announced a cost-cutting plan to reduce the size of their workforce by 10%. With a more efficient workforce, La Nibsa aims to boost profits by cutting down unnecessary spending

2. La Socovesa is planning to move their headquarters West to benefit from the hot market there, despite the fact that La Socovesa spent big money not-too-long-ago on renovating their current headquarters.

3. After 10 years of planning, designing, and testing the market, La Cinoca announced a new line of product.

4. After much consideration, La Venaseta has decided not to open up stores in Europe.

5. La Nibsa is shifting its focus to a new market segment because too much competition rushed into the field where La Nibsa was leading and drove down profits.

6. La Socovesa slashes prices in bid to become market's biggest player.

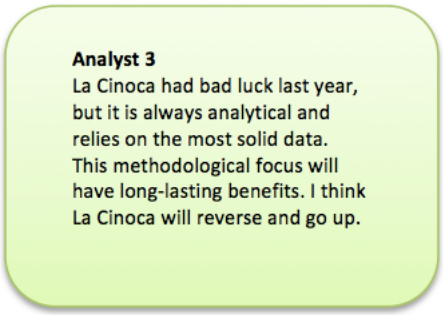
7. After trying to expand their own e-commerce website, La Cinoca has announced it will stop running its own website and partner with Amazon.com.

8. La Venaseta received a big tax break in the past year. Instead of using the money to expand the business, La Venaseta will use it to pay off loans and will be completely debt-free.

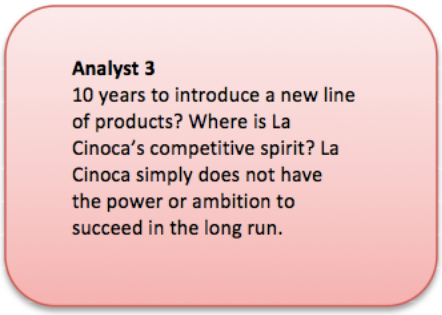
[34] These news excerpts are using female company names as an example. The male company names were: El Andacor, El Tecal, El Polledo, El Besalco

Examples of Comments

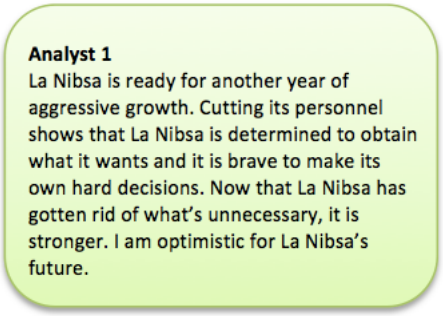
High Rationality



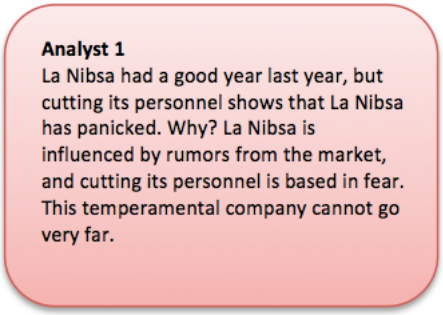
Low Strength



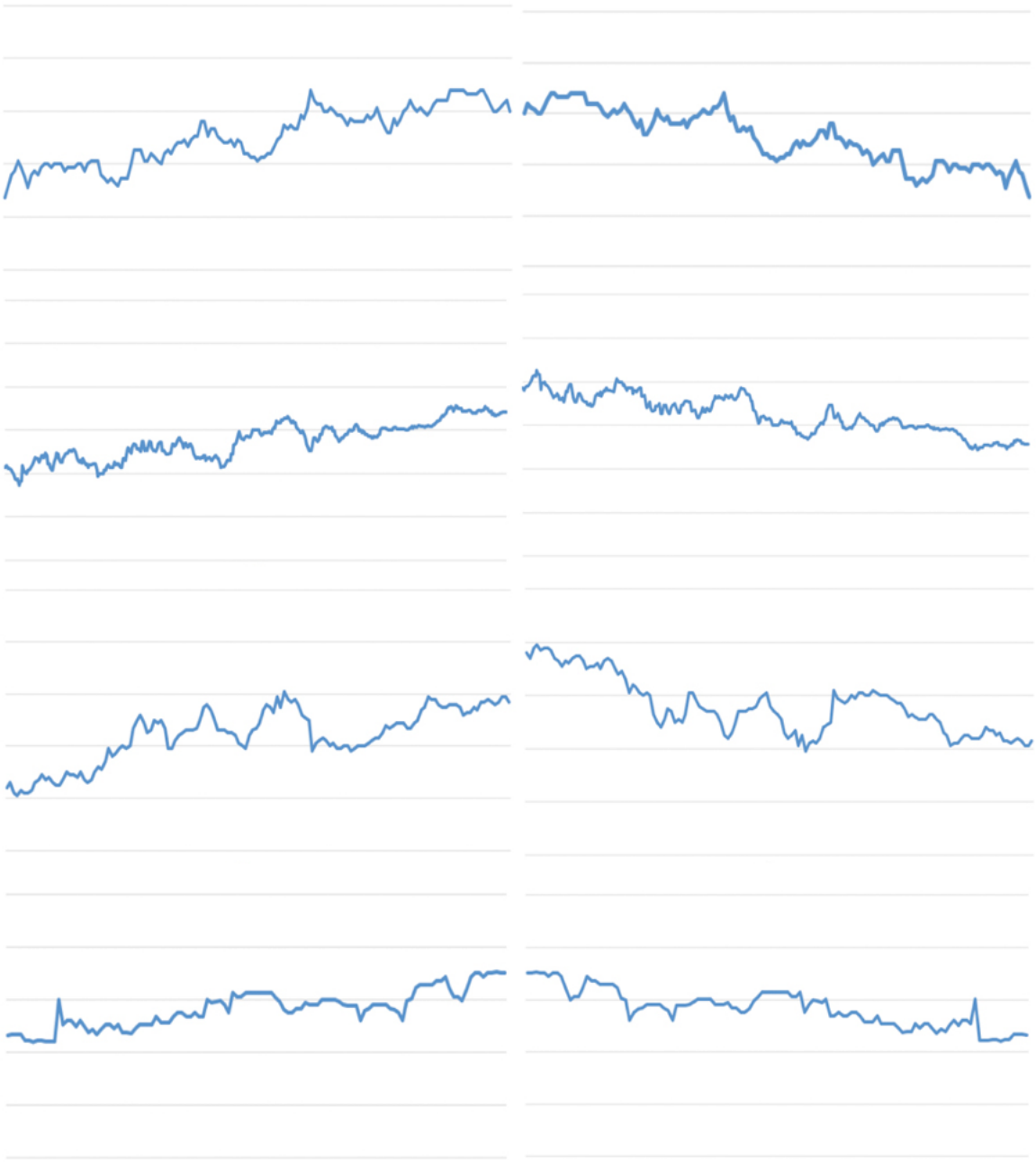
High Strength



Low Rationality



Graphs[35]



[35] The up and down graphs are mirror images of each other

APPENDIX C. PILOT STUDY 3.2

Top 20 nominated nouns for the animal Zhi (ranked by frequency)

noun	frequency
dog	17
cat	14
rabbit	12
bird	11
chicken	11
duck	9
mouse	8
hand	8
pig	6
lamb	6
Squirrel	6
shoe	5
cow	5
goose	4
tiger	4
worm	4
foot	4
turtle	3
frog	3

Top 20 nominated nouns for the object Zhi (ranked by frequency*)*

noun	frequency
pencil	11
pen	9
writing utensil (笔)	8
flower	7
stock	7
song	7
music	6
ball-point pen	6
brush	5
army	5
dance	4
cigarette	4
art marker (水彩笔)	3
marker (签字笔)	3
arrow	3
team	3
chopstick	2
flute	2
paper airplane	2

APPENDIX D. STUDY 3 EXAMPLES OF VERBS

Active expression - up:
领涨(leading the rise)，跑赢 (win in the race)，上攻(charge upwards)，冲高(sprint high)，正面强刚(to confront head on)，站上（ step on)，高歌猛进(march forward/upward aggressively, singing loud songs), 突破(break through)

Active expression - down:
砸(smash), 低走 (walk down), 跑输(lose in the race), 挣扎(struggle), 止步（ stop the steps)，跳水(dive)

Active expression - neutral/ambiguous direction:
坚守(to guard/maintain staunchly), 上蹿下跳 (jump up and down), 趴着不动(crouching there and not move)

Passive expression - up:
反弹(bounce back), 被顶(being pushed up), 得到支撑(being supported),被拉高 (being pulled up)

Passive expression - down:
回落(fall back), 下滑(slip down)，被封死(being completely blocked), 遇阻(encountered resistance)

Passive expression - neutral/ambiguous direction:
波动 (oscillate), 震荡(shake)

APPENDIX E. STUDY 4 PRIMING SENTENCES

Animal Primes

1. This cheetah (chicken) is looking for food and is not planning to share the food with that cheetah (chicken).

2. Those three wild wolves (squirrels) are by the tree far away from these two wild wolves (squirrels).

3. That bird is an eagle (a sparrow) with white feathers, different from the two eagles (sparrows) we saw yesterday.

4. Look at the two gorillas (flies) by the fresh meat. They are the two gorillas (flies) that the John has been looking for.

5. The tiger (hamster) in the bush had black stripes which helped it blend in to the environment. Consequently, the observers almost didn't realize that there was a tiger (hamster) in the bush.

Object Primes

1. That person bought two arrows (lipsticks) and placed one of the arrows (lipsticks) in the box.

2. The two arrows (lipsticks) were both red, but only the one in the box had feathers with stripes.

3. These five darts (two chopsticks) are made of steel, while those four darts (two chopsticks) are made of copper.

4. The man has two spears (pencils), and he prefers the heavy spear (light pencil) to the light spear (dark pencil).

5. Both spears (pencils) are sharp, but one of them was a gift from his dad.

Fillers

1. The TV was turned off by the last person that left the room.

2. The cup is full of lime-flavored sparkling water.

APPENDIX F. STUDY 1 DESCRIPIVE STATS FOR THE NEUTRAL CATEGORY

Note: results below are the averages across different lengths of holding and observation periods.

Outcome Variable	Calc Frequency	*N*	*M*	*SD*
overall momentum	day	61	.014	.038
	month	60	.018	.043
upward momentum	day	58	.025	.070
	month	57	.031	.073
downward momentum	day	61	.033	.052
	month	60	.032	.047