2025. Proc Ling Soc Amer 10(1). 5963. https://doi.org/10.3765/plsa.v10i1.5963.

Comparing form-based and meaning-based gender biases in pronoun resolution: Inferences from names and role nouns

Elsi Kaiser *

Abstract. Referring to people using only their name (e.g. *Jones came in*) is known to evoke the assumption that the individual is male. The same holds with some role nouns (e.g. *mechanic, boxer*). We explore these effects through the lens of pronoun interpretation in English. In two sentence-completion studies, we show that both form- (last-name-only style) and meaning-based gender biases (from role nouns) are powerful enough to eliminate otherwise robust verb semantic effects on pronoun interpretation (implicit causality). In addition, the results provide initial evidence that meaning-based biases (at least the ones tested here) can be stronger than form-based biases, which may stem from differences in the form-function mapping.

Keywords. pronoun interpretation; implicit causality; role nouns; gender stereotypes; reference; last names; experimental linguistics; sentence completion task

1. Introduction. Consider the classic riddle: *A father and his son are in a car accident. The father dies. The son is rushed to the hospital emergency room. The surgeon looks at the boy and says, "I can't operate on this boy. He's my son!" How can this be?* This riddle has been around for at least 50 years. According to Morehouse et al. (2022), it became widely known in the U.S. thanks to an episode of a TV series, *All in the family*, aired in 1972. However, when people are faced with this riddle for the first time, the possibility that the surgeon is the boy's mother is only infrequently mentioned. For example, Belle et al. (2021) found that only 30% of participants gave a 'mother' response. Similarly, Morehouse et al. (2022) found that only 27.1% of participants gave a 'mother' answer, whereas 47.4% gave a 'second father' answer (e.g., gay fathers, biological father). As Morehouse et al. note, these kinds of results "demonstrate the surprising strength of a gender occupational stereotype" (p.1).

In the present paper, we use pronoun resolution as a tool to tap into gender stereotypes associated with professions/role nouns (e.g. *surgeon, florist, boxer, nurse*) as well as certain kinds of referential forms, namely referring to someone with only their last name (family name, surname), as in *Smith walked in*, or *Jones was promoted*. This kind of last-name-only style is used more often for men than for women in U.S. English (e.g. McConnell-Ginet 2003, Atir & Ferguson 2018, Gardner & Brown-Schmidt 2024, Kaiser et al. 2022). We present two experiments that test the strength of the male bias associated with last-name-only style, assess whether it arises in informationally-impoverished contexts where participants do not have rich mental representations of the relevant referents, and test whether the male bias persists in the face of well-known verb semantic effects on reference resolution. The result show that yes, the bias does persist.

Furthermore, we provide an initial attempt at comparing the strength of the male bias associated with the last-name-only format relative to the bias strength of male-biased role nouns, in order to explore potential differences between biases stemming from words' lexical semantics and biases stemming from particular linguistic forms. The results suggest that biases rooted in

* Big thanks are due to Deborah Ho, Haley Hsu, Meona Khetrapal, Claire Benét Post, Madeline Rouse and Sabrina Soto for help with implementing the experiments and with data analysis. Thanks also are due to the University of Southern California Undergraduate Research Associates Program for funding, as well as the LSA 2025 audience for helpful feedback and comments. Author: Elsi Kaiser, University of Southern California (emkaiser@usc.edu)

lexical semantics can be stronger than biases stemming from particular linguistic forms, but we emphasize that this result is very preliminary and more work is needed on this topic.

1.1. REFERRING TO PEOPLE BY LAST-NAME ONLY. Whether a speaker chooses refer to someone by their first name, first and last name, a title followed by a last name, or last-name-only (or some other form) is influenced by numerous factors. The speaker's choices also influence others' impressions. This paper focuses on the phenomenon of referring to people with last-name only, as illustrated in (1-4). In (1), James Watson and Francis Crick are referred to by only their last names, whereas Rosalind Franklin is referred to using only her first name. In (2), Travis Kelce is referred to by last name only, whereas Taylor Swift is referred to using only her first name. In (3-4), we see examples of a male and a female professor referred to by last-name-only.

(1) I would go so far as to say that had Watson and Crick not come into Rosalind's photograph -- by hook or crook; whichever way it was -- they would have lost the race entirely (from a podcast by the Scientific American about Dr. Rosalind Franklin)

(2) Taylor is dating Kelce, the Kansas City Chiefs tight end. (Talking about Taylor Swift and Travis Kelce, CNBC)

(3) Johnson is a great professor. He is funny (from ratemyprofessor.com)

(4) Welsh is my favorite professor. She's just amazing (from reddit.com)

Last-name-only style does not carry an explicit marker of male gender (and does not have male phi-features, in formal terms) and can also be used for women (ex.4). However, it has been shown that, at least in U.S. English, in many contexts (e.g. politics, academia, sports, science, even informal conversation), men are more likely to be referred to by last-name-only than women (male bias, e.g. McConnell-Ginet 2003, Atir & Ferguson 2018, Gardner & Brown-Schmidt 2024, Kaiser et al. 2022, 2023, 2024a, 2024b). Moreover, referring to a researcher by last-name only results in them being judged more famous, more eminent, higher status and more deserving of awards (eminence bias, Atir & Ferguson 2018).

In their studies, Atir & Ferguson (2018) focused on naturalistic communication and other rich contexts where people knew a lot about each referent. For example, they used names of famous people or provided participants with information about scientists and their research. Furthermore, they did not specifically control the linguistic properties of the sentences mentioning the referents. While their work provides crucial naturalistic data, it does not aim to test specific linguistic properties related to pronoun resolution. In recent work, Gardner & Brown-Schmidt (2024) used a sentence-completion task with fragments such as *{Jordan/Smith/Jordan Smith} woke up early to walk the dog. After making coffee…* to test participants' assumptions about referent gender. Gardner & Brown-Schmidt (2024) found that when last-name-only format was used, participants "overwhelmingly used *he* to subsequently refer to the person, suggesting that participants inferred that the person was male" (p.1). Furthermore, in a memory study where participants read short stories, did some simple math problems, and answered questions about what they had read, Gardner & Brown-Schmidt found that when participants were asked about the gender of a person who had been introduced using last-name-only format, they also showed a strong skew towards male responses.

In order to take a closer look at different types of last names, in a recent paper Kaiser & Post (2025) compared last name subtypes in a production task. On each trial, participants saw a name accompanied by five bullet points of gender-neutral information about that person (e.g. *age: 31, went to school in: Savannah, Georgia, eye color: blue eyes*, and so on). We manipulated whether

the provided name was a gender-specific male or female first name (e.g. *Lucy, Greg*) or a last name (e.g. *Fields, Chapman, Atkinson, Hoffman*), and then analyzed what kinds of pronouns participants used when transforming the bullet points into complete sentences. In particular, we tested different kinds of last names to see if their morphological or semantic properties modulate the strength of the male bias: we tested last names ending in *-son* (e.g. *Atkinson*), ending in *-man* (e.g. *Hoffman)*, referring to stereotypically (historically) male professions (e.g. *Fisher*), based on words for male nobles (e.g. *Knight*), names referring to natural features/locations and landscape properties (e.g. *Fields*), and neutral last names lacking strong semantic links to common nouns (e.g. *Saunders*). Furthermore, we tested last names with components that could potentially be construed as having stereotypically female associations (e.g. flower terms, as in *Rosewood*).

The results from the Kaiser & Post (2025) experiment show that all last name types elicit a strong male bias – i.e., participants tend to use *he* pronouns when writing about the person – except for names with stereotypically female associations. More specifically, last names in the neutral, nature, *-man*, *-son,* nobility and profession conditions elicited more male than female pronouns (>50% he, <10% she), replicating the male bias found in prior work. In contrast, last names whose components can be thought of as having female associations elicit similar rates of *he* (26%) and *she* (31%) pronouns. We also looked at production of singular *they* to shed light on potential individual differences and language change, see Kaiser & Post (2025) for details.

In sum, prior work shows that last-name-only format, at least in U.S. English, has a strong male bias. In the present work, we test (1) how strongly this bias guides pronoun interpretation, compared to a well-known semantic bias (the implicit causality of verbs), and (2) compare the strength of this form-based gender bias to a meaning-based gender bias, namely gender biases associated with role nouns such as *surgeon* and *florist*.

1.2. MEANING-BASED VS. FORM-BASED EFFECTS. As mentioned above, one of the main aims of this work is to explore how form-based and meaning-based gender biases guide pronoun interpretation. With role nouns, people have expectations/beliefs about which roles are more often held by which gender(s) (e.g. Misersky et al. 2014). These expectations and beliefs are linked to a noun's lexical meaning. In other words, for a role noun to trigger a gender bias, a person needs to know the noun's meaning and to access noun's lexical entry. Thus, I refer to these as *meaning-based* effects.

In contrast, the male bias of last-name-only style is associated with a specific linguistic form – namely, referring to a person using a last name, as opposed to a first name or first and last name or some other option. I refer to this as a *form-based* effect. Thus, if someone hears or reads '*Smith walked in*' and assumes that Smith is male, this stems from the stylistic choice made by the speaker to use a last-name-only form. It is not associated with the content/meaning of a lexical entry, in contrast to a sentence such as '*The surgeon walked in*.' The two studies reported below aim to take initial steps to explore the strength of form- and meaning-based effects.

1.3. IMPLICIT CAUSALITY. Another key aim of the present work is to explore how the strength of the gender biases of last-name-only format and role nouns stack up against another well-known factor that guides pronoun interpretation, namely the phenomenon of implicit causality. It is well-known that verb semantics influence pronoun interpretation. Usually the term 'implicit causality' is used specifically for contexts like those in (5-6), where the second clause provides an explanation of the event or state in the first clause, as signaled by the connective *because*. Verbs differ in how strong of an expectation they create for the following explanation to focus on the preceding subject or object. Some verbs (e.g. *impress, annoy*) tend to elicit subject interpretations of subsequent pronouns, as in (5): these are known as IC1 verbs. Others (e.g. *admire,*

criticize) tend to elicit object interpretations of subsequent pronouns; these are called IC2 verbs.

(5) Andy impressed Eric because he…. (IC1 verb, favors subject)

(6) Andy admired Eric because he…. (IC2 verb, favors object)

These patterns are well-established in prior work (e.g. Caramazza et al. 1977, McKoon et al. 1993, Rudolph & Försterling 1997, Koornneef & van Berkum 2006, Ferstl et al. 2011, Hartshorne & Snedeker 2013, Bott & Solstad 2014, Hartshorne et al. 2015, see also Patterson et al. 2022 for related work). Given the stability of implicit causality (IC) effects on pronoun interpretation, IC effects are well-suited for testing the strength of the male bias associated with last-name-only format. In a context such as *{Jones/The surgeon} impressed Lisa because she…* , how do people interpret the pronoun *she*? The verb bias of *impress* strongly biases an interpretation where *she* refers to the subject *Jones* or *the surgeon*, but can this bias be erased by the male bias of last-name-only style or the role noun?

2. Experiment 1: Last-name only format

2.1. METHOD

2.1.1. PARTICIPANTS.

Participation occurred study remotely over the internet. In total 101 people participated, and 7 were excluded for not being U.S.-born native English speakers, one for reporting impaired vision (which can impact language input) and two for failing a pre-experiment attention check. This left 91 adult native U.S. English speaking participants for the final analysis (53 female, 36 male, 1 prefer not to say, 1 other).

2.1.2. MATERIALS AND DESIGN.

To assess the strength of the male bias of last-name-only format, we constructed two-clause items as illustrated in (7). The two clauses are linked by the connective *because*, and the second clause is truncated after the prompt pronoun (*he* or *she*). The first clause is a transitive and contains an implicit causality verb. The last-name-only referent is in subject position (as in 7a,b) or in object position (as in 7c,d), depending on the implicit causality of the verb, as explained below. The other argument is a first name whose typical gender matches the pronoun's gender (e.g. Eric, Frank when the prompt pronoun is *he*; Amanda, Claire when the prompt pronoun is *she*).

Furthermore, we also manipulated whether the implicit causality verb is biased towards the subject (IC1) or the object (IC2) – in other words, whether the semantics of the IC verb pushes the pronoun to favor the subject (as in 7a,b) or the object (as in 7c,d). We tested both IC1 and IC2 verbs in order to control for effects of grammatical role, given prior work suggesting that subjects are privileged as antecedents of pronouns (and thus IC effects may be stronger with IC1 verbs than IC2 verbs). So, we tested both IC1 and IC2 verbs to ensure that our results generalize beyond a particular grammatical role.

(7) Example item
- (a) *IC1 verb (semantics favor the subject) + he*
  Smith impressed Eric because he
- (b) *IC1 verb (semantics favor the subject) + she*
  Smith impressed Amanda because she
- (c) *IC2 verb (semantics favor the subject) + he*
  Frank promoted Mayfield because he
- (d) IC2 verb (semantics favor the subject) + she
  Claire promoted Mayfield because she

Crucially, because the last-name-only referent is always in the position favored by the verb semantics, this set-up allows us to test whether participants' pronoun interpretation is guided by verb semantics: In other words, how often do people interpret the prompt pronoun as referring to the last-name-only referent that is favored by verb semantics? As we discuss more in the predictions section, the key conditions of interest are the ones where the prompt pronoun is *she*. Do participants interpret *she* as referring to the last-name-only referent (e.g. *Smith*, *Mayfield*), in line with the verb bias? Or is the male bias of the last-name-only format so strong that participants tend to interpret *she* as referring to the gender-matching first name, even though that name does *not* match the verb's implicit causality bias?

The implicit causality verbs were selected based on norms collected by Hartshorne & Snedeker (2013) and Ferstl et al (2011), to ensure that the IC1 and IC2 verbs are reliably subject-biased and object-biased respectively. In the target stimuli, we used 20 IC1 verbs and 20 IC2 verbs, balanced for positive and negative polarity (e.g. *amaze, admire* vs. *infuriate, criticize*). The average subject preference of the IC1 verbs was 74.63% and the average object preference of the IC2 verbs was 78.46% (based on combined norming data from both Hartshorne & Snedeker 2013 and Ferstl et al. 2011).[1] The mean preference strength for both verb groups combined (i.e., subject preference for IC1 verbs and object preference for IC2 verbs) was 76.55%.

For the first names, we avoided gender-neutral names (but see Kaiser et al. 2024a similar for work using gender-neutral names) and instead used first names which have clear typical gender associations and would be familiar to U.S. English speakers (e.g. *Claire, Eric*). For the last names, we largely used names with origins in Great Britain and Ireland, e.g. *Miller, Watkins, Jenkins, O'Connor, Wright, Hawkins*, and avoided names ending in *-son* or *-man*, to avoid potentially strengthening the male bias. (For work testing whether semantic and morphological properties of a name modulate the male bias, see Kaiser & Post 2025. We are currently conducting follow-ups using last names with more diverse geographical and linguistic origins, which requires norming to ensure people recognize a last name as a last name, instead of a first name.)

In addition to 20 target items, the study also included 22 fillers. Fillers also consisted of an initial clause and a partial second clause which contained a mix of syntactic structures and a variety of first and last names.

2.1.3. PROCEDURE AND DATA ANALYSIS. The study was conducted over the internet using Qualtrics (Provo, UT). Participants were instructed to write a natural-sounding completion for each sentence fragment. We opted for a sentence-completion task because it allows us to tap into people's pronoun interpretation patterns, which will provide information about how they interpret last-name-only format, *without* us having to ask any explicit questions about what gender people think the last-name-only referent has.

Each item was shown on a separate screen, and participants completed the task at their own pace. Examples of completions are provided in Table 1. After data collection, the continuations were double-coded by two coders blind to the conditions: the names and pronouns were removed for purposes of coding and replaced with the placeholder labels 'subject,' 'object,' and 'pronoun' – in other words, the coders only saw things like '*subject impressed object because pronoun had done a great job with the project.'* This was done to ensure that coders' own potential biases do

[1] Further analyses (incorporating norms from both Hartshorne & Snedeker 2013 and Ferstl et al. 2011) confirm that the sets of IC1 and IC2 verbs do not differ in the strength of their verb biases (unpaired t-test, $t(38)=1.18$, $p>0.2$): IC1 verbs prefer the subject as strongly as IC2 verbs prefer the object.

not distort the data. Completions were coded for whether the pronoun prompt refers to the preceding subject, object, or whether this is unclear. (5.8% of the data were coded as unclear.)

Sentence fragment	*Continuation written by participant*	*Coded as*
Smith impressed Eric because he	… had done a great job with the project.	he=subject
Smith impressed Amanda because she	… thought his performance was great.	she=object
Claire promoted Mayfield because she	… had impressed Claire with her skills.	she=object

Table 1. Example continuations provided by participants in Experiment 1

An additional question has to do with the question of whether there are additional anaphoric references after the subject-position prompt pronoun. In this study, we did not yet analyze whether the other referent from the preceding clause was mentioned later on in the continuation or what form was used for this, but see Song & Kaiser (2024) for recent work indicating that the referential properties of a subject pronoun in two-pronoun sentences can differ from those of a subject pronoun in one0pronoun sentences (e.g. *she helped her* vs. *she helped Kate*), even in non-parallelism contexts where the clauses involve an explanation relation indicated by *because*.

2.2. PREDICTIONS. The design of Experiment 1 pits the male bias of last-name-only format against verbs' implicit causality, and thus allows us to see which has a stronger effect on reference resolution. Crucially, the last-name-only referent is always in the position favored by verb semantics (in other words, it is the referent favored by the verb's implicit causality bias). Thus, if pronoun interpretation is guided by implicit causality, as shown in a large body of prior work, participants should interpret the prompt pronoun as referring to the preceding subject with IC1 verbs, and the preceding object with IC2, regardless of whether the pronoun is *he* or *she*. However, if last-name-only format has a male bias – i.e., if participants construe a last-name-only referent as male – this will clash with the IC bias and could weaken or eliminate the IC effects.

Thus, if only implicit causality matters in pronoun interpretation, we should find above-chance rates of subject interpretations with IC1 verbs and above-chance rates of object interpretations with IC2 verbs. In other words, we should find high rates of *verb-bias-compatible interpretations* regardless of pronoun gender. So, because the last-name referent is in the position favored by verb bias, this means we should observe high rates of last-name interpretations with both *he* and *she* pronouns in all four conditions.

However, if last-name-only format has a male bias, we expect to find a decrease in the rate of last-name interpretations when the prompt pronoun is *she*. In other words, the *IC1 + she* and *IC2 + she* conditions are predicted to show fewer verb-bias-compatible continuations than the *IC1 + he* and *IC2 + he* conditions.

3. Experiment 1 Results and discussion: Last-name only. Figure 1 shows how often participants interpret the prompt pronoun as referring to the last-name-only referent, as a function of verb type (IC1, IC2) and pronoun gender (*he/she*). [2] The data are presented in terms of the proportion of verb-bias-compatible continuations: for IC1 verbs, how often the pronoun is interpreted as referring to the preceding *subject* and for IC2 verbs, how often the pronoun is interpreted as referring to the preceding *object*. Plotting and analyzing the data in terms of verb-

[2] In Figures 1 and 2, positive and negative verbs are collapsed. While there are some differences in the pronoun interpretation patterns after positive and negative verbs, they are not relevant for the main claims being made in this paper, so I do not discuss them in detail here.

bias compatibility, instead of grammatical role, allows us to present the data for IC1 and IC2 verbs on the same dimension.

The two bars on the left show the results for conditions where the prompt pronoun was *he.* The fact that both bars are tall shows that participants overwhelmingly interpret *he* as referring to the verb-bias compatible last-name-only referent (>80%, above chance, p's<.001 using *glmer* intercept-only models). In other words, participants interpret the pronoun in the manner predicted by the verb's semantic IC bias. This strong effect of verb semantics is in line with prior work on implicit causality effects (see Section 1).

However, a different picture emerges when the prompt pronoun is *she,* in the two bars on the right. These bars hover around 50%. This shows that participants interpret *she* as referring to the last-name-only referent at much lower rates (40%-50%; no conditions are significantly above chance): essentially, participants are 'avoiding' using *she* for last-name-only antecedents. Thus, although verb bias clearly favors the last-name-only referent (e.g. *Smith impressed Amanda because she*), when the pronoun is *she,* participants interpret *Smith* as the antecedent of the pronoun at most 50% of the time.

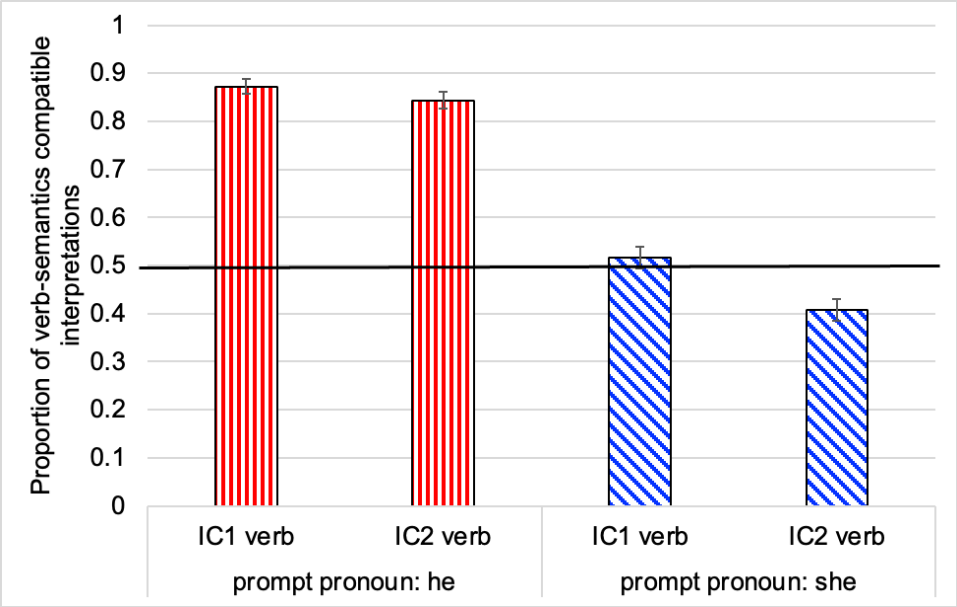

Figure 1. Experiment 1 Last-name-only format: Proportion of verb-bias-compatible pronoun interpretations (i.e., how often is the prompt pronoun interpreted as referring to the preceding subject with IC1 verbs and preceding object with IC2 verbs).

In sum, these results show that last-name-only format has a robust male bias, which arises even in a very informationally-impoverished experimental context. Furthermore, the male bias is strong enough that it is able to able to overcome strong semantic biases on pronoun resolution: We find no significant preference for the preceding subject or object with *she*, despite the verbs having a strong implicit causality bias that favors one referent over the other.

Although we do not include an analysis of participants' gender here, it's worth noting that that both male and female participants exhibit a male bias with last-name-only format.

In addition to testing last-name-only format, we have also conducted similar experiments on last-name only format with hyphenated last names (e.g. *Smith-Jones*), gender-neutral first

names and last names preceded by the title Dr. (e.g. Kaiser et al. 2022, 2023, 2024a, 2024b). Furthermore, in Kaiser et al. (2023) we compare humans' biases with those of ChatGPT.

So far, we have seen strong evidence for form-based effects. In other words, the last-name-only form is – at least for U.S. English speakers – strongly associated with a male default. In the next section, we turn to effects of meaning-based factors, namely the gender biases associated with role nouns, to see how they compare.

4. Experiment 2: Male-biased role nouns

4.1. METHOD.

4.1.1. PARTICIPANTS. Participation took place the same way as in Experiment 1. In total 60 people participated, and 4 were excluded for not being U.S.-born native English speakers. This left 56 adult native U.S. English speaking participants for the final analysis (38 female, 18 male).

4.1.2. MATERIALS AND DESIGN. The design was the same as in Experiment 1, but instead of using 20 different last names, we used twenty different male-biased role nouns, selected using gender norms from Misersky et al. (2014). An example item is in (8).

In their norming study, Misersky and colleagues asked 1,408 participants to estimate the ratio of men and women in different social and occupational groups on an 11-point rating scale ranging from 0 % women and 100 % men to 100 % women and 0 % men. The role nouns selected for Experiment 2 had a mean rating of 76% men (range 65-84, SD 4.2) – in other words, when the norming participants were asked to indicate what proportion of the people who occupy that role are men, participants' average response was 76%. Put another way, on average, for the roles we tested: approximately three out of four people who have a certain role are men (3/4 = 75%).

This number can be contextualized relative to existing work on last-name-only format: In an unconstrained production study, Atir & Ferguson (2018, Experiment 4) asked participants to turn bullet points about a scientist named Dolores Berson or Douglas Berson into full sentences (see also Kaiser & Post 2025). Atir & Ferguson found that participants were approximately four times more likely to refer to the male scientist using last-name-only (25% of references) than to use last-name-only for the female scientist (6% of references). (This experiment did not test other role nouns.) Putting it differently, proportionally speaking, approximately three out of four last-name-only references are to a man (25% with Douglas vs. 6% with Dolores).[3]

Although this is admittedly a very rough comparison, it suggests that the male bias of last-name-only format can perhaps be viewed as roughly comparable to the strength of the male bias of the role nouns that we tested. However, it's clear that questions remain about the baseline strength of these two biases, and more work is needed on this, in order to gain a more direct measure of how strongly last-name-only format is associated with male vs. female referents.

(8) Example item
  (a) *IC1 verb (semantics favor the subject) + he*
  The guard impressed Eric because he
  (b) *IC1 verb (semantics favor the subject) + she*
  The guard impressed Amanda because she

[3] Atir & Ferguson report percentages, not raw numbers, but because theirs was an experimental study, we can assume that the number of participants who wrote about Dolores Berson and Douglas Berson respectively were (approximately) the same, so the percentages of last-name-only format can be compared in a meaningful way – at least for the purposes of the rough comparison we are aiming for here.

(c) *IC2 verb (semantics favor the subject) + he*
Frank promoted the engineer because he
(d) *IC2 verb (semantics favor the subject) + she*
Claire promoted the engineer because she

Like Experiment 1, this study also included 22 fillers in addition to the 20 targets. The fillers are similar to Experiment 1, but were revised to include some mentions of gender-neutral role nouns (e.g. *novelist, tour guide*), to ensure that role nouns occurred in both target and filler items.

4.1.3. PROCEDURE AND DATA ANALYSIS. The procedure for data collection was the same as in Experiment 1. Examples of continuations are in Table 2. As in Experiment 1, the continuations were double-coded by two coders blind to the conditions. Just like Experiment 1, to minimize effects of coders' own biases, the names, role nouns and prompt pronouns were removed for coding and replaced with the placeholder labels 'subject,' 'object,' and 'pronoun' – in other words, the coders only saw things like '*subject impressed object because pronoun never moved from his post.'* Completions were coded for whether the pronoun prompt refers to the preceding subject, object, or whether this is unclear. (3.8% of the data were coded as unclear.)

Sentence fragment	*Continuation written by participant*	*Coded as*
The guard impressed Eric because he	… never moved from his post.	he=subject
The guard impressed Amanda because she	… thought he would be able to recover her lost purse.	she=object
Claire promoted the engineer because she	…. had been going above and beyond in her work.	she=object

Table 2. Example continuations provided by participants in Experiment 2

4.2 PREDICTIONS. The predictions have the same logic as Experiment 1. Here, we pit the male bias of the role nouns against verbs' implicit causality, to test which has a stronger effect on pronoun interpretation. Echoing the logic of Experiment 1, the male-biased role noun is in the position favored by verb semantics (by the verb's implicit causality bias). If pronoun interpretation is guided only by implicit causality, participants should show a strong preference to interpret the pronoun as referring to the verb-bias-compatible antecedent, regardless of whether the pronoun is *he* or *she*. But if the male-bias of the role nouns make participants less likely to interpret *she* as referring to those role nouns – i.e., people tend not to use *she* for *the engineer*, even though this is of course entirely possible – then this will clash with the IC bias. In this case, the *IC1 + she* and *IC2 + she* conditions are predicted to show fewer verb-bias-compatible continuations than the *IC1 + he* and *IC2 + he* conditions, echoing the logic of Experiment 1.

Thus, independent of Experiment 1, Experiment 2 allows us to test the relative strength of the gender bias of role nouns vs. IC verbs on pronoun interpretation, to see what happens when they clash: Is one more influential than the other? Recall that the mean preference strength for the IC1 and IC2 verbs (i.e., subject preference for IC1 verbs and object preference for IC2 verbs) was 76.55%. Furthermore, recall that the mean men male bias of the role nouns that we used was 76%, as discussed above. Although it's clear that these numbers are *not* an apples-to-apples comparison and it would be great to have more nuanced measures, it will be interesting to see what happens when these two effects are pitted against each other.

In what follows, we first consider the results for Experiment 2 on its own and then compare them to the results of Experiment 1.

5. Experiment 2 Results and discussion: Male-biased role nouns. Figure 2 shows how often people interpret the pronoun as referring to the male-biased role noun. The figure again shows the proportion of verb-bias-compatible continuations (IC1 verbs, how often the pronoun refers to the preceding *subject*; for IC2 verbs, how often it refers to the preceding *object*).

In Figure 2, the two leftmost bars, where the prompt pronoun is *he*, show that participants overwhelmingly interpret *he* as referring to the referent favored by the verb bias, i.e., the male-biased role noun (>70%, above chance, p's<.01). When the prompt pronoun is *she*, we again see a different pattern: The two bars on the right are much lower. This shows that participants interpret *she* as referring to the male-biased role noun at much lower rates (<40%, below chance, p's<.05). In other words, participants are 'reluctant' to interpret she as referring to role nouns like *engineer, roofer, sailor, electrician*, even though the verb bias favors this referent (e.g. *The electrician inspired Sarah because she…*).'

It's also worth noting that In Experiment 2, with both *he* and *she*, we see a higher rate of verb-bias-compatible interpretations with IC1 than IC2 verbs; we attribute this to the favored antecedent being the subject with IC1 verbs, and subjects are independently known to be privileged antecedents for subject-position pronouns. (Kaiser et al. 2022, 2023, 2024 also found hints of this subjecthood effect in other last-name-only studies.) Thus, this is not unexpected and it is orthogonal to the main focus of this work.

Furthermore, although we do not report an analysis of participants' gender here, our results indicate that both male and female participants interpret *she* as referring to male-biased role nouns less often than *he*.

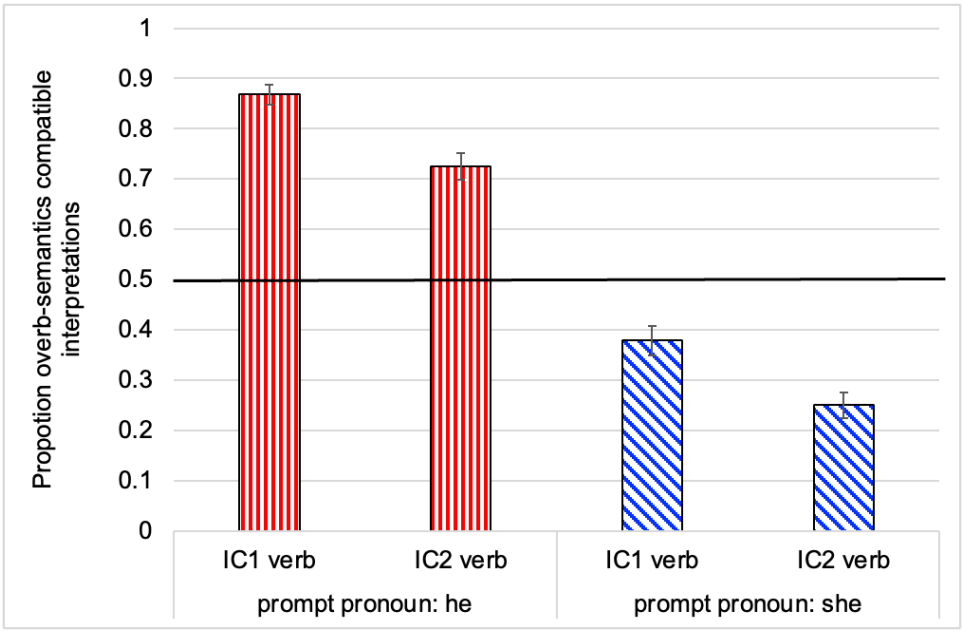

Figure 2. Experiment 2 Male-biased role nouns: Proportion of verb-bias-compatible pronoun interpretations (i.e., how often is the prompt pronoun interpreted as referring to the preceding subject with IC1 verbs and preceding object with IC2 verbs).

5.2. COMPARING EXPERIMENTS 1 AND 2. We also statistically compared the results of Experiments 1 and 2. The conditions with *he* pronouns do not differ significantly in Experiment 1 vs. Experiment 2 in terms of the proportion of verb-compatible continuations (glmer, z=1.93, p=0.054), though we see hints of a marginal difference. In contrast, the conditions with *she* pro-

nouns differ in the two experiments. Specifically, *she* is significantly likely to be interpreted as referring to male-biased role nouns (Experiment 2) than to last-name-only referents (Experiment 1, glmer, $z=3.25$, $p<0.01$). This suggests that while both last-name-only format and male-biased role nouns both exhibit a male bias, the effects of this bias on pronoun interpretation appear to be stronger with male-biased role nouns than with last-name only, at least for the role nouns that were tested in the present work. This could be taken as a possible indication that meaning-based effects are stronger than form-based effects.

6. General Discussion. The present paper uses pronoun resolution as a tool to tap into gender stereotypes associated with role nouns (e.g. *surgeon, florist, boxer, nurse*) and last-name-only format (e.g. *Jones made dinner)*. The two sentence-completion experiments reported in this paper show that last-name-only format (Experiment 1) and male-biased role nouns (Experiment 2) both have significant effects on pronoun interpretation. In particular, the male default associated with these last-name-only and male biased role nouns is strong enough to complete with robust and widely-observed effects of verb implicit causality.

Specifically, even when the implicit causality of a verb strongly biases a pronoun to pick out a particular referent (e.g. the subject with an IC1 verb as in *{Smith/The guard} impressed Amanda because she…*), participants are at chance or below chance in interpreting a pronoun as referring to that referent when the gender of the pronoun mismatches the male bias of last-name-only format or the male bias of the role noun. This shows that the male bias of last-name-only format – at least in the U.S. English context – is strong enough to overcome strong IC biases, even in simple experimental contexts where participants are not familiar with and have minimal information about the referents being talked about (cf. Atir & Ferguson 2018).

Furthermore, the differences between Experiments 1 and 2 suggest that form-based differences may be weaker than meaning-based differences: The gender biases associated with role nouns appear to have a stronger effect that the male bias of last-name-only format. However, it's important to emphasize that further work is needed on this topic, in particular a better way of measuring the strength of the male bias of last-name-only format. Even though the present work tried to test role nouns whose male bias is in some sense 'comparable' to the male bias of last-name-only (see Section 4), these comparisons are very rough and more systematic assessments are needed. Thus, this conclusion is only preliminary.

However, that being said, the idea that meaning-based biases stemming from role nouns can be stronger than form-based biases receives support the nature of these phenomena. One factor to consider has to do with the notion of arbitrariness. Form-based effects – i.e. the fact that last-name-only tends to be associated with male reference, at least in the U.S. – is arguably fairly arbitrary. In essence, we are dealing with a social convention, presumably related to the fact that women have typically changed their last name upon marriage (at least in many parts of the western world). Crucially, one could imagine an alternative situation where last-name-only reference is associated with another property of the referent, for example age: One could imagine a world where older people who have had a certain last name for longer are referred to by last-name-only more often than younger people, regardless of gender. Given that the mapping between form (last-name-only) and its sociolinguistic function (male) is at least somewhat arbitrary, it is perhaps unsurprising that the gender bias of last-name-only is weaker than that of role nouns.

Furthermore, the gender bias of role nouns may also be strengthened by the essentializing properties of nouns. Although the studies reported here did not directly compare role nouns and adjectives (but see Kaiser & Adji 2025), it is interesting to note that researchers have argued that

nouns tend to favor essentialist thinking and stereotypical inferences more than adjectives. For example, cognitive psychology research suggests that "When a noun rather than an adjective is used, both children and adults draw more robust inferences and judge features to be more inheritable, persistent, and explanatory" (Ritchie 2021, p.471, see also Gelman & Markman (1986), Markman (1989), Markman & Smith (cited by Markman 1989), Carnaghi et al. (2008) and others). These earlier studies did not focus on role nouns or gender bias, but the findings align with the idea that the gender biases of role nouns can be robust. For example, Carnaghi et al. (2008) found that nouns elicit stronger essentializing inferences than adjectives in contexts like *Mark is an athlete/athletic*. They conclude that "nouns have a greater likelihood than adjectives to induce stereotype-congruent expectancies" (Carnaghi et al. 2008, p.846). These essentializing properties of nouns seem to align well with the idea that male-biased role nouns could have a stronger effect on the interpretation of gender-marked pronouns than last-name-only format.

It also bears emphasizing that the current work focuses on the U.S. context and U.S. English. Differences in referential patterns across languages and social contexts are an important direction for future work. Ultimately, many questions remain open for future investigation.

References

Atir, Stav & Melissa J. Ferguson. 2018. How gender determines the way we speak about professionals. *Proceedings of the National Academy of Sciences* 115(28). 7278-7283.

Belle, Deborah, Ashley B. Tartarilla, Mikaela Wapman, Marisa Schlieber & Andrea E. Mercurio. 2021. "I Can't Operate, that Boy Is my Son!": Gender Schemas and a Classic Riddle. *Sex Roles* 85. 161-171.

Bott, Oliver & Torgrim Solstad. 2014. From verbs to discourse: A novel account of implicit causality. In Barbara Hemforth, Barbara Mertins & Cathrine Fabricius-Hansen (eds), *Psycholinguistic approaches to meaning and understanding across languages,* 213-251. Springer.

Caramazza, Alfonso, Ellen Grober, Catherine Garvey & Jack Yates. 1977. Comprehension of anaphoric pronouns. *Journal of Verbal Learning and Verbal Behavior* 16(5). 601-609.

Carnaghi, Andrea, Anne Maass, Sara Gresta, Mauro Bianchi, Mara Cadinu & Luciano Arcuri. 2008. Nomina sunt omina: on the inductive potential of nouns and adjectives in person perception. *Journal of personality and social psychology* 94(5). 839-859.

Ferstl, Evelyn C., Alan Garnham, & Christina Manouilidou. 2011. Implicit causality bias in English: A corpus of 300 verbs. *Behavior Research Methods* 43. 124-135.

Gardner, Bethany, Sarah Brown-Schmidt & Sarah Brown-Schmidt. 2024. Biased inferences about gender from names. *Glossa Psycholinguistics* 3(1). https://escholarship.org/uc/item/0zm3m3hh

Gelman, Susan A. & Ellen M. Markman, Ellen. 1986. Categories and induction in young children. *Cognition* 23. 183-209.

Hartshorne, Joshua K. & Jesse Snedeker. 2013. Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes* 28. 1474-1508.

Hartshorne, Joshua K., Timothy J. O'Donnell & Joshua B. Tenenbaum. 2015. The causes and consequences explicit in verbs. *Language, Cognition and Neuroscience* 30(6). 716-734.

Kaiser, Elsi & Ashley Adji. 2025. On gender stereotypicality in nouns and adjectives: Comparing humans, large language models and text-to-image generators. *Proceedings of the Linguistic Society of America* 10(1). 5954-5954. https://doi.org/10.3765/plsa.v10i1.5954

Kaiser, Elsi & Claire Benét Post. 2025. Names and Pronouns with and without Gender Features: A Production Study of Singular *they*. In Nikolas Webster et al. (eds.), *Proceedings of the 41st West Coast Conference on Formal Linguistics (WCCFL),* 270-277. Somerville, MA: Cascadilla Proceedings Project. https://www.lingref.com/cpp/wccfl/41/paper3755.pdf

Kaiser, Elsi, Deborah Ho, Haley Hsu, Claire Bénet Post, Claire & Madeline Rouse, Madeline. 2022. Referring to someone using only their last name: Insights from gender-marked pronouns. Talk at *Linguistic Evidence 2022*. Paris, France: Université de Paris – CNRS, France.

Kaiser, Elsi, Claire Benét Post, Deborah Ho, Haley Hsu & Madeline Rouse Madeline. 2023. Gender biases, pronouns and artificial intelligence: Comparing humans and ChatGPT. Talk at *Linguistic Intersections of Language and Gender*. Düsseldorf, Germany: Heinrich-Heine-Universität.

Kaiser, Elsi, Claire Benét Post, Deborah Ho, Haley Hsu, Meona Khetrapal, Madeline Rouse & Sabrina Soto. 2024a. Social information in reference resolution: Last names and gender bias. Poster at the *37th Annual Conference on Human Sentence Processing (HSP 2024)*. Ann Arbor MI: University of Michigan.

Kaiser, Elsi, Meona Khetrapal & Sabrina Soto. 2024b. Last names, hyphens and gender-neutral first names: Testing assumptions about gender. Talk at *the 58th Annual Meeting of the Canadian Society for the Study of Names / Société canadienne d'onomastique*. Montréal, Canada: Université de Montréal.

Koornneef, Arnout W. & Jos JA Van Berkum. 2006. On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language* 54, 445-465.

Markman, Ellen M. 1989. *Categorization and naming in children*. Cambridge, MA: MIT Press.

McConnell-Ginet, Sally. 2003. "What's in a name?" Social labeling and gender practices. In Janet Holmes & Miriam Meyerhoff (eds.), *Handbook of Language and Gender,* 69-97. Oxford: Blackwell.

McKoon, Gail, Steven B. Greene & Roger Ratcliff. 1993. Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19(5). 1040-1052.

Misersky, Julia, Pascal M. Gygax, Paolo Canal, Ute Gabriel, Alan Garnham, Friederike Braun, Tania Chiarini, Kjellrun Englund, Adriana Hanulikova, Anton Öttl, Jana Valdrova, Lisa Von Stockhausen & Sabine Sczesny. 2014. Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods* 46. 841-871.

Morehouse, Kirsten N., Benedek Kurdi, Ece Hakim & Mahzarin R. Banaji. 2022. When a stereotype dumbfounds: Probing the nature of the surgeon= male belief. *Current Research in Ecological and Social Psychology* 3: 100044.

Patterson, Clare, Petra B. Schumacher, Bruno Nicenboim, Johannes Hagen & Andrew Kehler. 2022. A Bayesian approach to German personal and demonstrative pronouns. *Frontiers in Psychology* 12. 672927. https://doi.org/10.3389/fpsyg.2021.672927

Ritchie, Katherine. 2021. Essentializing language and the prospects for ameliorative projects. *Ethics*, 131. 3460-488

Rudolph, Udo, and Friedrich Försterling. 1997. The psychological causality implicit in verbs: A review. *Psychological Bulletin* 121(2). 192-218.

Song, Jina & Elsi Kaiser. 2024. Effects of referential structure on pronoun interpretation. *Language, Cognition and Neuroscience* 39(1): 98-117.