Generalization in Legal Argumentation

Frank Zenker\textsuperscript{a,b,c}, Christian Dahlman\textsuperscript{d}, Sverker Sikström\textsuperscript{e}, Lena Wahlberg\textsuperscript{d}, and Farhan Sarwar\textsuperscript{e}

\textsuperscript{a}Department of Philosophy, Lund University, Lund, Sweden; \textsuperscript{b}Department of Philosophy, Boğaziçi University, 34342 Istanbul, Turkey; \textsuperscript{c}Int. Center for Formal Ontology, Warsaw University of Technology, Poland; \textsuperscript{d}Faculty of Law, Lund University, Lund, Sweden; \textsuperscript{e}Department of Psychology, Lund University, Lund, Sweden

**ABSTRACT**

When interpreting a natural language argument that generalizes over a contextually relevant category, audiences are likely to activate the category prototype and transfer its characteristics onto category instances. A generalized argument can thus appear more (respectively less) persuasive than one mentioning a specific category instance, provided the argument’s claim is more (less) warranted for the prototype than for the instance (positive and negative prototype effect). To investigate this effect in legal contexts using mock-scenarios, professional and lay judges at Swedish courts evaluated the persuasiveness of arguments giving a generalized or a specific description of an eyewitness. The generalized version described the witness either as an alcohol-intoxicated person or as a child, while the specific version varied both the amount of alcohol consumed (two vs. five glasses of wine) and the child’s age (four vs. 12 years). To investigate the effect of legal expertise on argument selection, moreover, law and social science students evaluate the persuasiveness of both argument versions. Though we observed statistically significant prototype effects as well as expertise effects, results were mixed and sometimes ran counter to normative expectation.

**KEYWORDS**

Argumentation; decision-making; evidence; expertise effect; generalization; lay judge; legal context; persuasiveness; professional judge; prototype effect

**Introduction**

Assume that right before observing a crime, an eyewitness drinks a glass of wine. To doubt her reliability before a court, further assume that a defense lawyer claims: “an alcohol-intoxicated person is not a reliable witness.” This generalization argument (Inman & Baron, 1996; Krumm & Corning, 2008) conveys the (allegedly self-evident) claim that an intoxicated person is never a reliable witness. The persuasive strategy here is to suggest that all instances of the category INTOXICATED PERSON share the same characteristic – namely being unreliable as a witness.

Drinking four glasses (rather than one glass) of wine, moreover, makes one a more typical instance of the category INTOXICATED PERSON, thus bringing one closer to the category’s prototype. Prototypes generally serve
reasoners in interpreting visual, spoken, or written information quickly (Mervis & Rosch, 1981), because the prototype fits a category’s characteristics better than non-prototypical instances (Cantor & Mischel, 1979; Inman & Baron, 1996; Krumm & Corning, 2008; McCaughey & Strohmer, 2005; Rosch, 1975). Prior research suggests that activating the prototype can “over-shadow” the category instance (Dahlman, Sarwar, Bååth, Wahlberg, & Sikström, 2016; Lakoff, 1987, p. 85). Audiences who evaluate an argument describing a category instance would thus transfer their evaluation of the contextually relevant prototype to the instance.

Where prototype-features replace instance-features – such that one evaluates a case as if it had the prototype’s rather than its actual features – the literature speaks of a radical prototype effect (Winter, 1988, p. 1386) or an assimilation-to-prototype-effect (Winter, 2001, pp. 150–151). To explain the differential persuasiveness of generalized vs. specific arguments, then, one may postulate a prototype effect (Lakoff, 1987; Rosch, 1975). For instance, if the prototypical intoxicated person drinks four glasses of wine, then audiences may evaluate an argument applying the category INTOXICATED PERSON to someone having consumed two glasses as if she had consumed two additional glasses.

Audiences may therefore agree more strongly with the focal reliability claim, for they would (rightly) find the claim more warranted that someone who consumed four glasses of wine is a comparatively less reliable witness, than the claim that someone who consumed two glasses of wine is a comparatively less reliable witness. We refer to this as a positive prototype effect. In the opposite direction, audiences may agree less strongly should the prototype overshadow the instance in ways that make the reliability claim less warranted for the prototype than for the actual case. For instance, if an argument applying the category INTOXICATED PERSON activates the prototypical four glasses of wine, then a someone having in fact consumed five glasses of wine is judged as if she had consumed one glass fewer. We refer to this a negative prototype effect.

The study investigates the persuasiveness of generalization arguments in legal contexts using a sample of professional and lay judges at Swedish courts, and a sample of law and social science students at Lund University, Sweden. Our (directional) hypothesis on how prototype effects influence argument persuasiveness allows for three predictions: (i) if an audience finds a focal claim on average more warranted for the prototype than for the instance, then a generalization argument lets it agree with the claim on average more strongly (positive prototype effect). (ii) If an audience finds the focal claim on average less warranted for the prototype than for the actual case, then a generalization argument lets it agree on average less strongly with it (negative prototype effect). (iii) If an audience finds the focal claim on average equally warranted for the
prototype as for the actual case, then a generalization argument has on average no effect on argument persuasiveness.

**Background**

**Generalization in argumentation**

Categories normally surface as linguistic generalizations of the form “(All) subject-term [copula] predicate-term”. To evaluate instances of this form, audiences regularly defer to features of the category prototype (Inman & Baron, 1996; Krumm & Corning, 2008). The prototype’s features are thought to block the interpretation of features that are distinct to the category instance – replacing more specific with less specific information. Thus, activating the category-prototype rather than the category-instance may (partially) explain the differential persuasiveness of generalization arguments (Nathan & Lord, 1983).

The shift between instance and prototype is known as a radical prototype effect (Winter, 1988, p. 1386), a halo evaluation effect, or simply a bias toward the prototype (Feldman, 1981; Nathan & Lord, 1983). The problematic inference is that all category members allegedly share the same characteristics, for instance that “all intoxicated people are unreliable witnesses, irrespective of the amount of alcohol they consume”. To count as a member of a focal category, however, instances must (often) satisfy a necessary condition. An alcohol-intoxicated witness, for instance, must have consumed a minimum amount that serves as a contextually defined boundary quantity. Ignoring this quantity when subsuming an instance under a category results in a category transgression (Dahlman et al., 2016), insofar as the “intoxicated person”-prototype exists only above the boundary quantity. The descriptive baseline for a meaningful comparison is the case where audiences find the argument’s generalized version as persuasive as its specific version.

Normatively, if the category-instance falls below the boundary quantity, then the corresponding argument should be less persuasive than if the instance sat above the boundary. For instance, let the boundary quantity be \(\frac{1}{4} \times x\) glasses of wine, and let the prototypical quantity be \(x\) glasses. A category transgression can lead to a positive prototype effect, given that audiences rate the generalization argument – “the witness’s reliability is reduced because she was alcohol-intoxicated” – as more persuasive than the specific argument “the witness’s reliability is reduced because she consumed \(x < 4\) glasses of wine.” (Normatively, persuasiveness-ratings should monotonously correlate with \(x\).)

By contrast, a negative prototype effect occurs if audiences rate the argument’s specific version as more persuasive than its generalized version.

Since both effects depend on how audiences understand the category prototype, the effect size may differ across individuals, cultures, or contexts.
(Fehr, 1988; Horowitz & Turan, 2008; Hovland and Sherif, 1980; Martz, Strohmer, Fitzgerald, Daniel, & Arm, 2009).

**Professional vs. Lay judges**

Swedish courts comprise professional as well as lay judges, who each have one vote, and whose number depends on the court’s level and a case’s nature (The Swedish Code of Judicial Procedure, 1942, p. 70). Professional judges have received legal training, command experience in handling a variety of legal arguments, and are normally employed for life. Lay judges, by contrast, come from diverse professional backgrounds, serve a fixed term of four years, and tend to represent various political orientations (Diesen, 2001).

Insofar as professional judges are better than lay judges at “overrid[ing] their intuitive reactions with complex, deliberative thought,” they are thought to be better equipped at handling complex legal arguments (Guthrie, Rachlinski, & Wistrich, 2007, p. 9; Pettersson, Dahlman, & Sarwar, 2018). Conversely, scholars have raised concerns that lay judges are more likely to base their judgments on intuition and emotion (Diesen, 2001; Goldbach & Hans, 2014). Both groups of judges therefore probably differ in how persuasive they find a generalization argument, a difference that counts as soon as one group has the majority.

**This study**

**Set up**

In experiment 1, we presented both professional and lay judges with a mock-scenario conveying alcohol consumption as a sufficient reason to doubt an eyewitness’s reliability. Judges read the scenario in one of two randomly assigned versions. The generalized version described the witness either as an alcohol-intoxicated person or as a child; the specific versions varied both the amount of alcohol consumed (two vs. five glasses of wine) and the child’s age (four vs. 12 years). Participants were instructed to rate the argument’s persuasiveness. Experiment 2 presented law and social science students with the same mock-scenario, but each participant read both argument versions (generalized and specific). Participants were instructed to select the version they would present in court to convince a judge or jury.

The purpose of this set-up is to shed light upon the question whether positive or negative persuasive prototype effects occur in the two scenarios. The main question in the age scenario was: if the witness is in fact 12 years old, is the argument “the witness is less reliable because of her age” on average more persuasive if the witness is described as a “child”, as opposed a “twelve year old” (positive prototype effect)? Respectively, if the witness is in
fact four years old, is the argument on average less persuasive when the
witness is described as a “child” (negative prototype effect)? Accordingly,
the main question in the wine scenario was: if the witness had consumed
only two glasses of wine, is the argument questioning her reliability as a
witness on average more persuasive if the witness is described as “intoxicated”
(positive prototype effect)? Respectively, if the witness had consumed five
glasses of wine, is the argument on average less persuasive if the witness is
described as “intoxicated” (negative prototype effect)?

For these questions, four hypotheses spell out the expected responses.

**Hypotheses**

**Hypothesis 1: positive prototype effect**

If the argument’s specific version presents the focal quantity \(Q_F\) as falling between the boundary quantity \(Q_B\) and the prototypical quantity \(Q_P\), then both professional and lay judges should find the generalized argument \(A_G\) on average more persuasive (PER) than the specific argument \(A_S\). Hypothesis 1 thus states: if \(Q_B > Q_F > Q_P\), then \(\text{PER}(A_G) > \text{PER}(A_S)\). For instance, if a child’s prototypical age is seven years and the mock-scenario specifies the eyewitness’s age as 12 years, then one should expect a positive prototype effect for the CHILD prototype. The same holds for the ALCOHOL-INTOXICATED PERSON prototype, assuming the prototypical amount is four glasses of wine and the eyewitness’ consumed amount is two glasses. After all, a higher persuasiveness rating is warranted if a seven-year-old child has consumed four glasses of wine, compared to a twelve-year-old child having consumed two glasses. In each case, the effect grounds in the difference between the prototypical properties (seven years; four glasses) and the scenario’s specific quantities.

**Hypothesis 2: negative prototype effect**

If the argument’s specific version presents the focal quantity as falling above the prototypical quantity, then one should expect that both professional and lay judges find the specific version on average more persuasive than its generalized version. Hypothesis 2 thus states: if \(Q_P > Q_F\), then \(\text{PER}(A_S) > \text{PER}(A_G)\). Provided a child’s prototypical age is seven years, and the scenario specifies the eyewitness’s age as four years, one should expect a negative prototype effect for the CHILD prototype. By parity of reasoning, the same holds for the ALCOHOL-INTOXICATED PERSON prototype, if the prototypical amount is four glasses of wine and consumed alcohol amount is given as five glasses.

**Hypothesis 3: expertise effect in persuasiveness-ratings**

Evaluating an argument’s persuasiveness properly requires both training and experience (Goldbach & Hans, 2014). A difference in expertise thus entails that professional judges should be on average better at this than lay judges.
One should therefore expect that positive and negative prototype effects are stronger among lay than among professional judges.

**Hypothesis 4: expertise effect in argument selection**

Similarly, one should expect (at least traces of) this effect size-difference when comparing law with social science students. After all, law students receive training in selecting and presenting persuasive legal arguments (Clements, 2013; Toye, 2013), while social science students normally lack such training. Thus, if the prototype compares with the mock-scenario such that the argument’s generalized version is on average comparatively more warranted, then one should expect that law students select this version on average more often than social science students.

**Experiment 1**

**Participants**

To collect data from professional judges, we cooperated with the chief judge at each Swedish municipal court, who forwarded to them our questionnaires and return envelopes. To collect data from lay judges, we sent them questionnaires and return envelopes directly. Each questionnaire contained ten scenarios, with the focal age and the wine scenario at the 4th and 8th position, respectively. The other eight scenarios were filler items (addressing questions unrelated to this study). We awaited replies for about four months, and during that time sent all judges a reminder. (A few late replies were not included in the analysis.)

In this way, we collected data from 668 participants (283 professional judges and 385 lay judges); 324 were female (professional = 126, lay = 198) and 344 male (professional = 157, lay = 187). Of the 667 questionnaires sent to professional judges, 283 were returned (response rate ≈42%); all were used in our analysis. To lay judges, we sent out 738 questionnaires individually, of which 391 were returned (response rate ≈53%). After excluding six incomplete ones, we used 385 questionnaires in our analysis.

**Design**

We used a 2x2x2-between subject design with two scenarios:

1. **Age scenario:** the first factor was the judge’s position (professional vs. lay). The second factor was the witness’ age (four vs. 12 years). The third factor was the argument version (specific vs. generalized).
2. **Wine scenario:** the first factor was the judge’s position (professional vs. lay). The second factor was the alcohol amount the witness had consumed (two vs. five glasses of wine). The third factor was the argument version (specific vs. generalized).
Materials

Participants were exposed to both the age and the wine scenario, as developed and used previously by Dahlman et al. (2016), presented in Swedish. (We give an English translation below).

Age scenario

Michael J is accused of manslaughter. According to the prosecutor’s description, using a knife he stabbed Kevin O to death in a garbage room in Södertälje [a Swedish city]. The prosecutor’s main witness is Emily E. She is $x$ (four or twelve) years old, and saw Kevin O enter the garbage room, accompanied by a man wearing a black hoodie. Emily E identifies Michael J as the man in the black hoodie.

After reading the scenario, each participant read either the argument’s specific or its generalized version. The specific version was: “confidence in Emily E’s testimony is adversely affected by the fact that she is under $x$ years” ($x = 5$ or $x = 13$). The generalized version: “confidence in Emily E’s testimony is adversely affected by the fact that she is a child.” Participants were instructed to read the scenario first, and then indicate to which extent they agree with the argument version on a 9-point scale, ranging from 1 (strongly disagree) to 9 (strongly agree). We randomly varied both the child’s age (four vs. 12 years) and the argument version (specific vs. generalized).

Wine scenario

“Daniel A was charged with drug dealing. According to the prosecutor’s description, he had sold drugs in a club in Västerås [a Swedish city]. One of the witnesses in the trial is Carina J, who saw Daniel A while she was sitting with a friend at the bar having wine. During Carina J’s cross-examination it emerges that, when she made her observations, she had consumed $x$ glasses of wine” ($x=2$ or $x=5$).

The argument’s specific version was: “confidence in Carina J’s testimony is adversely affected by the fact that she had consumed $x$ glasses of wine when she made her observations” ($x = 2$ or $x = 5$). The generalized version was: “confidence in Carina J’s testimony is adversely affected by the fact that she was intoxicated with alcohol when she made her observations.” Again, after participants read the scenario, they were presented either with the argument’s specific or its generalized version. Participants were then asked to indicate to which extent they agree with the argument about the eyewitness’s reliability, using a 9-point scale ranging from 1 (strongly disagree) to 9 (strongly agree). The amount of wine consumed (two or five glasses) and the argument version (specific or generalized) were randomized.
Results

In the age scenario, we did not observe statistically significant prototype effects. In the wine scenario, we observed a positive prototype effect and a negative prototype effect for lay judges in line with Hypothesis 1 and Hypothesis 2. When the witness had consumed two glasses of wine, then lay judges ($t(203) = -2.77, p < .01$) agreed on average significantly more strongly with the argument’s generalized version, which described the witness as “intoxicated” (mean = 3.75, SD = 2.20), than with the version stating the consumed amount as “two glasses” (mean = 2.96, SD = 1.90). Prototype-theory explains this result as a positive prototype effect (see Figure 2). By contrast, when the case’s circumstances stated that the witness had consumed five glasses of wine, then lay judges ($t(176) = 2.08, p < .04$) agreed significantly more strongly with the argument’s specific version, stating the consumed amount as “five glasses” (mean = 6.29, SD = 2.05), than with the version that used the generalization “intoxicated” (mean = 5.65, SD = 2.09). This result is explained as a negative prototype effect (see Figure 3).

Age scenario

We analyzed data using a 2x2x2 between subjects ANOVA. The first factor was the judge’s position (professional vs. lay), the second factor was the eyewitness’s age (four vs. 12 years), the third factor was the argument version (specific vs. generalized). Results showed a significant main effect for the judge’s position: compared to lay judges, professional judges tended to agree on average more strongly with the argument ($F(1,667) = 29.15, p = .001$, partial $\eta^2 = .04$).

To further probe the main effect, we conducted four $t$-tests (with a Bonferroni correction) using a significance level of $p = .01$. When the eyewitness’ age was specified as four years, professional judges agreed on average significantly more strongly than lay judges with both the argument’s specific version ($t(153) = 3.90, p < .0001$; Cohen’s $d = 0.63$) and its generalized version ($t(158) = 3.14, p < .002$; Cohen’s $d = 0.51$). When the eyewitness’ age was 12 years, moreover, professional judges agreed on average significantly more strongly than lay judges with the argument’s generalized version ($t(176) = 3.02, p = .003$; Cohen’s $d = 0.45$). Both groups of judges did not differ significantly in agreeing with the argument’s specific version (see Table 1 and Figure 1).

Wine scenario

We analyzed data for the wine scenario using a 2x2x2 between subjects ANOVA. The first factor was the judge’s position (professional vs. lay judge), the second factor was the amount of alcohol consumed (two vs. five glasses of wine), the third factor was the argument version (specific vs. generalized). Results showed a significant three-way interaction between
To further probe the three-way interaction, we conducted four $2 \times 2$ ANOVAs. Two of these addressed the two vs. five glasses of wine the eyewitness had consumed, where the first factor was the judge’s position (professional vs. lay), and the second factor was the argument version (specific vs. generalized). For the two glasses of wine version, we observed a significant main effect of the argument version on participants’ persuasiveness ratings ($F(1,351) = 6.19, p = .01; \text{partial } \eta^2 = .02$, observed power = .70). Participants agreed on average significantly more strongly with the argument’s
generalized (mean = 3.74, SD = 2.15) than with its specific version (mean = 3.14, SD = 2.02). We thus observed a positive prototype effect.

To probe the main effect for the argument version separately for professional and lay judges, we conducted two independent sample t-tests. Lay judges agreed on average significantly more strongly with the argument’s generalized (mean = 3.75, SD = 2.20) than with its specific version (mean = 2.96, SD = 1.90), \(t(203) = -2.77, p < .01\). Compared to lay judges, professional judges also agreed on average somewhat more strongly more with the argument’s generalized (mean = 3.71, SD = 2.10) than with its specific version (mean = 3.39, SD = 2.16). But this difference was statistically insignificant \((t(144) = -0.94, p < .35)\).

Results for the argument’s five glasses of wine-version showed a significant interaction effect between the judge’s position and the argument version (\(F(1,312) = 11.38, p = .001\); partial \(\eta^2 = .04\), observed power = .92). To probe this interaction effect further, we conducted four independent sample t-tests. The first two t-tests compared lay and professional judges with respect to agreeing to the argument’s generalized and its specific version; the other two

**Figure 2.** Professional and lay judges’ average agreement to the argument’s specific and generalized version, given the amount of alcohol consumed.
-tests compared each group of judges separately with respect to agreeing to one or the other argument version.

Professional judges (mean = 6.87, SD = 2.14) agreed on average significantly more strongly with the argument’s generalized version than lay judges (mean = 5.65, SD = 2.09), $t(152) = 3.57, p < .0001$, while professional and lay judges did not differ significantly in agreeing to its specific version.

Table 2. ANOVA test-results: means and standard deviations (in brackets) reporting participants’ agreement given the alcohol amount and the type of argument ($n = 668$).

| Judges’ position | Alcohol amount: 2 glasses | Alcohol amount: 5 glasses |
|------------------|----------------------------|----------------------------|
|                  | Specific argument | Generalized argument | Specific argument | Generalized argument |
| Professional     | 3.38 (2.16)       | 3.71 (2.09)           | 5.86 (2.37)       | 6.87 (2.14)           |
| Lay              | 2.96 (1.89)       | 3.75 (2.20)           | 6.29 (2.05)       | 5.65 (2.09)           |
| Professional and lay | 3.13 (2.02) | 3.73 (2.15)           | 6.12 (2.19)       | 6.21 (2.19)           |

Figure 3. Lay and professional judges’ average agreement to the argument’s specific and generalized version.
Moreover, professional judges agreed on average significantly more strongly with the argument’s generalized (mean = 6.87, SD = 2.14) than with its specific version (mean = 5.86, SD = 2.37), $t(132) = −2.60, p < .01$. We thus observed a positive prototype effect. By contrast, lay judges ($t(176) = 2.08, p < .04$) agreed on average significantly more strongly with the argument’s specific (mean = 6.29, SD = 2.05) than its generalized version (mean = 5.65, SD = 2.09). We thus observed a negative prototype effect.

The third and fourth ANOVA addressed the status of lay and professional judges. The first factor was the amount of alcohol consumed (two vs. five glasses of wine), and the second factor was the argument version (specific vs. generalized). For lay judges, we observed a significant interaction effect between the amount of alcohol consumed and the argument version ($F(1,383) = 11.58, p = .001; \text{partial } \eta^2 = .03; \text{observed power} = .92$). To probe this interaction effect further, we conducted two $t$-tests. If the witness was described as having consumed two glasses of wine, lay judges ($t(203) = −2.77, p < .01$) agreed on average significantly more strongly with the argument’s generalized (mean = 3.75, SD = 2.20) than with its specific version (mean = 2.96, SD = 1.90). We thus observed a positive prototype effect. By contrast, if the witness was described as having consumed five glasses of wine, lay judges ($t(176) = 2.08, p < .04$) agreed on average significantly more strongly with the argument’s specific (mean = 6.29, SD = 2.05) than with its generalized version (mean = 5.65, SD = 2.09). We thus observed a negative prototype effect (see Figure 3).

For professional judges, we observed a significant main effect of the argument version on persuasiveness ratings ($F(1,280) = 6.59, p = .01; \text{partial } \eta^2 = .02, \text{observed power} = .73$). Professional judges agreed on average significantly more strongly with the argument’s generalized (mean = 5.28, SD = 2.64) than with its specific version (mean = 4.52, SD = 2.57). To probe this main effect further, we conducted two independent sample $t$-tests: one addressed the two glasses of wine scenario, the other the five glasses of wine scenario. If the witness was described as having consumed five glasses of wine, professional judges (mean = 6.87, SD = 2.14) agreed on average significantly more strongly with the argument’s generalized than with its specific version (mean = 5.86, SD = 2.37), $t(132) = −2.60, p < .01$. By contrast, if the witness was described as having consumed two glasses of wine, professional judges’ average agreement did not differ markedly between the generalized and the specific version.

We also observed a significant main effect of the argument version on persuasiveness ratings ($F(1,1663) = 5.03, p = .03; \text{partial } \eta^2 = .01, \text{observed power} = .61$). All participants agreed on average significantly more strongly with the argument’s generalized (mean = 4.90, SD = 2.50) than with its specific version (mean = 4.54, SD = 2.57). Finally, we observed a significant main effect of the amount of alcohol on persuasiveness ratings ($F(1,1663) = 265.64$, $p < .01$).
All participants agreed on average significantly more strongly with the argument’s specific version if the witness was described as having consumed five (mean = 6.17, SD = 2.19) as opposed to two glasses of wine (mean = 3.43, SD = 2.10).

**Experiment 2**

**Legal training**

Complex argumentation generally is difficult to evaluate, the main reason why lawyers receive extensive training in argumentation (Clements, 2013; Goldbach & Hans, 2014; Toye, 2013). To investigate whether legal training compares positively to social science training – by raising awareness for the persuasive effect of using generalizations – we asked law and social science students to identify the argument version (generalized vs. specific) they would rather select in order to convince a judge and jury.

**Participants, design, and materials**

Our sample comprised 160 students at Lund University, Sweden. 101 of them studied at the Faculty of Law, and 59 at the Faculty of Social Sciences. All students were contacted on campus, and were handed a printed questionnaire (plus consent form) on which they recorded their responses.

Our between-subjects design had two conditions reflecting the participants’ field of study (Faculty of Law vs. Faculty of Social Science). The design reused the age scenario from experiment 1, except that the argument’s specific version presented the eyewitness exclusively as being 12 years old.

Participants read both the specific and the generalized argument version according to these instructions:

> You are Williams’ lawyer and can use one of the following two arguments to question the reliability of the witness. Which one would you choose to convince the judge and the jury?

**Results**

On average, law students preferred the argument’s generalized version (64.8%) over its specific version (35.2%), while social science students did not differ markedly between preferring the generalized (50.8%) and the specific version (49.2%). A chi-square (goodness of fit) test indeed showed a significant difference among law students ($X^2(1,N = 105) = 9.12, p < .002$), but not among social science students ($X^2(1,N = 59) = .02, p < .89$; see Table 3).
Discussion

Positive prototype effect

As per hypothesis 1, if the focal quantity falls between the boundary quantity and the prototypical quantity, then both professional and lay judges should evaluate the argument’s generalized version as on average more persuasive than its specific version (positive prototype effect), because this evaluation is more warranted for the prototype than for the actual case. In the age scenario, we thus expected a positive prototype effect from using the term “child” when the eyewitness was 12 years old, assuming the prototypical age of a child is around seven years. In the wine scenario, we similarly expected a positive prototype effect from using the terms “intoxicated with alcohol” if the eyewitness had consumed two glasses of wine, assuming the prototypical alcohol amount an intoxicated person consumes is around four glasses of wine. Both predictions are based on the earlier finding, that using a generalization makes an argument more persuasive if the judgment is more warranted for the prototype than for the instance (Dahlman et al., 2016).

In the wine scenario, if the witness was described as having consumed two glasses of wine, results are consistent with this prediction for lay judges. They agreed on average more strongly with the argument’s generalized than its specific version, thus displaying a positive prototype effect. This is in line with previous empirical results suggesting that audiences generalize a category’s prototypical features to an instance (Inman & Baron, 1996; Krumm & Corning, 2008). By contrast, professional judges did on average not differ significantly between agreeing to the argument’s generalized or its specific version. To explain this, one may speculate that judges found their judgment equally warranted for the prototype as for the instance (Dahlman et al., 2016).

In the age scenario, neither the interaction effect nor the main effect of argument version onto persuasiveness rating were statistically significant. We therefore did not further probe a positive prototype effect for professional or lay judges. As the means in Figure 1 show, however, if the witness’ age was set to 12 years, then professional judges nevertheless agreed on average considerably more strongly with the argument’s generalized (mean = 4.43, SD = 2.34) than its specific version (mean = 3.67, SD = 2.31), thus displaying a positive prototype effect. This difference may well have been significant, had

Table 3. Number and percentage of law and social science students (n = 106) who preferred the argument’s generalized over its specific version.

| Condition            | Generalized argument | Specific argument |
|----------------------|-----------------------|-------------------|
| Law students         | 68 (64.8%)            | 37 (35.2%)        |
| Social science students | 30 (51.5%)            | 29 (49.5%)        |
it not been “buried” under the comparisons of the mixed ANOVA’s grand means. Consistent with this speculation, an independent sample t-test shows that professional judges did indeed agree on average significantly more strongly with the argument’s generalized version than its specific version ($t(146) = -1.98, p < .05$). Yet the effect size (Cohen’s $d = .01$) was very small.

**Negative prototype effect**

As per hypothesis 2, if a specific case falls above the prototype, then professional and lay judges should evaluate the argument’s specific version as being on average more persuasive than its generalized version (*negative prototype effect*), because this judgment is more warranted for the actual case than for the prototype. In the age scenario, we thus expected a negative prototype effect from using the term “child” if the eyewitness was described as four years old, assuming a child’s prototypical age is around seven years. In the wine scenario, we similarly expected a negative prototype effect from using the terms “intoxicated with alcohol” if the witness had consumed five glasses of wine, assuming the prototypical amount an intoxicated person consumes is around four glasses of wine.

Consistent with this hypothesis, if the consumed amount was set to five glasses, then lay judges agreed on average more strongly with the argument’s specific than with its generalized version, displaying a *negative prototype effect*. Inconsistent with hypothesis 2, by contrast, professional judges instead agreed on average more strongly with the argument’s generalized than with its specific version, displaying a *positive prototype effect* (Figure 2). What may explain this difference is the distinct amount of alcohol that professional and lay judges associated to the prototype INTOXICATED PERSON (Martz et al., 2009). Here, one potential explanatory factor may be *political stance* (e.g., being an active party member, or not), insofar as lay judges seek to behave politically correct, and insofar as alcohol consumption remains politically contested in Sweden. Professional judges, by contrast, might be less motivated to display such behavior and may rather associate the Swedish population’s general understanding of this amount. Similar considerations, at any rate, suffice to motivate associating different alcohol amounts to the term “intoxicated”.

Regarding hypothesis 1 and 2, our mixed results suggest that a generalization can make an argument appear more persuasive, *only if* participants find their judgment more warranted for the prototype than for a specific case.

---

1For lay judges, the prototypical alcohol amount may be less than five glasses (perhaps four glasses, as per hypothesis 2). Lay judges would thus find their judgment more warranted for the specific case than for the prototype (*negative prototype effect*). For professional judges, by contrast, the prototypical amount may be more than five glasses (perhaps six?). They would thus find their judgment more warranted for the prototype than for specific case (*positive prototype effect*).
case. This holds insofar as judges share a similar understanding of the category’s prototype mentioned in the generalized argument. As predicted, for instance, lay judges displayed a positive prototype effect if the alcohol amount was set to two glasses, but a negative prototype effect if it was set to five glasses.

Of course, no law of nature could force judges to share a similar understanding of the prototypical amount. Persuasiveness ratings can therefore very well differ from our predictions. For instance, if the alcohol amount was set to five glasses, we did – contrary to expecting a negative prototype effect – in fact observe a positive effect.

**Expertise effect in judging argument strength**

Results generally suggest that professional judges are more likely than lay judges to find their judgments warranted for the prototype. Thus, differences in training and experience may not suffice to protect from false judgments in precisely those cases where a positive prototype effect would amount to a judgment error.

Because understanding complex argumentation properly requires training, hypothesis 3 predicts that professional judges do on average arrive at more accurate persuasiveness rating than lay judges, who lack such training (Goldbach & Hans, 2014). (Here, “more accurate” means that a judgment better traces a case’s facts.) Thus, if hypothesis 1 predicted a positive prototype effect, then we expected a stronger effect among lay than among professional judges. Similarly, if hypothesis 2 predicted a negative prototype effect, then we expected a stronger negative prototype effect for lay than for professional judges.

Contrary to expectation, however, results differed in direction for both the age and the wine scenario. In the age scenario, professional judges did on average agree more strongly with a given argument version than lay judges. To explain this pattern, one may assume that professional judges employ a more “generous” intoxication-criterion (Lindholm, 2008), on one hand, and also refer to a general overconfidence among legal professionals, on the other (Goodman-Delahunty, Granhag, Hartwig, & Loftus, 2010). Although lay judges are regularly criticized for their non-technical role in the legal system (e.g., Diesen, 2001), if they can be held responsible, then they may nevertheless arrive at more careful judgments (Tetlock, 1983). Our results could be an instance of this pattern.

In the age scenario, moreover, if the eyewitness’ age was set to 12 years, then professional judges did on average agree significantly more strongly than lay judges with the argument’s generalized version. But there was no difference regarding the specific version. We rather expected that lay judges would instead agree on average more strongly than professional judges with
the generalized version. Yet, lay judges in our sample again made more careful judgments, whereas their professional colleagues acted generously, and perhaps overconfident, too. Finally, if the alcohol amount was set to two glasses, both professional and lay judges did not differ markedly in agreeing to the generalized or the specific version. This suggests that two glasses of wine may be a lower “intoxication boundary.”

Results are consistent with predicting that lay judges show a larger negative prototype effect, as they agree on average less strongly than professional judges with the argument’s generalized version. We nevertheless expected that professional judges would also show a negative prototype effect – if a smaller effect than among lay judges. Instead, professional judges displayed a positive prototype effect, indicating a possible difference in how they understand the prototypical amount.

In the context of a negative prototype effect, moreover, if the witness age was set to four years, then lay judges made poorer (i.e., less accurate) judgments than their professional colleagues. This held for both scenarios. If the alcohol amount was set to five glasses, then both groups of judges did on average commit judgment errors: lay judges were affected by a negative prototype effect, and professional judges by a positive prototype effect. This underlines the importance of future research with similar scenarios and prototypes among relevant groups.

Most of the significant results emerged in the wine scenario, and were consistent with hypotheses 1 and 2, yet partially inconsistent with hypothesis 3. One potential explanatory factor refers to differences between both scenarios, as judges’ understanding of the respective prototypical amounts can vary with context (Horowitz & Turan, 2008) and groups (Martz et al., 2009). Another such factor is mental fatigue (Wright et al., 2007). The wine scenario appeared at the questionnaire’s end, at the eighth of ten positions. This may have impaired performance, causing (more) judgment errors to arise as participants became (more) tired (Danziger, Levav, & Avnaim-Pessoa, 2011). Yet more research is needed to properly address both factors.

**Expertise effect in argument choice**

For cases where a judgment is more warranted for the prototype than for the instance, hypothesis 4 predicted that, compared to social science students, law students choose the argument’s generalized version on average more frequently than its specific version. This prediction pivots on law students’ training in legal argumentation having comparatively improved their argumentation skills (Clements, 2013; Toye, 2013). Consistent with this hypothesis, the majority of law students preferred the argument’s generalized version, while preferences were split (50:50) among social science students. This suggests that law students understand persuasiveness differences more
readily if the argument’s specific version seems weak, a pattern that their training explains readily (Clements, 2013; Toye, 2013).

**General discussion**

The use of linguistic generalizations in legal argumentation potentially affects judges’ and jurors’ degree of agreement with an argument, even though the case’s facts are the same. This may lead to over- or underestimating the evidence’s true value (positive and negative prototype effect), and thus to unjust verdicts (Kalai, 1993). Unlike in case of leading questions, after all, courts have so far placed no restrictions on similar manipulations.

Regarding professional competence, experiment 2 showed that training in legal argumentation may plausibly have helped law students to identify the argument’s generalized version as being comparatively more persuasive. Although the professional judges in our sample presumably received similar training, they nevertheless broadly failed to appreciate the risk of agreeing more strongly with the argument’s generalized version.

**Conclusions**

Some of the results presented here are consistent with the general assumption that lay judges are more prone to be affected by generalizations than professional judges. We observed positive as well as negative prototype effects among lay judges, but not among professional judges. We moreover observed expertise effects among law and social science students. Though we expected prototype effects in line with hypothesis 1 and hypothesis 2 among professional judges, we did not observe such effects. Since the observed effects sizes were small to medium, moreover, future research should supply additional samples. This would help to estimate the true effect size, and thus clarify whether effects perhaps arise from our scenarios’ lack of ecological validity, or in fact mark actual between group-differences.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

Research was supported by the Swedish Research Council (Vetenskapsrådet) and the Ragnar Söderberg Foundation (Ragnar Söderbergs Stiftelser).
References

Cantor, N., & Mischel, W. (1979). Prototypes in person perception. In B. Leonard (Ed.), *Advances in experimental social psychology* (Vol. 12, pp. 3–52). New York, NY: Academic Press.

Clements, C. S. (2013). Perception and persuasion in legal argumentation: Using informal fallacies and cognitive biases to win the war of words. *Brigham Young University Law Review, 2013*(2), 319–361.

Cohen, C. E. (1981). Person categories and social perception: Testing some boundaries of the processing effect of prior knowledge. *Journal of Personality and Social Psychology, 40*(3), 441–452. doi:10.1037/0022-3514.40.3.441

Dahlman, C., Sarwar, F., Bååth, R., Wahlberg, L., & Sikström, S. (2016). Prototype effect and the persuasiveness of generalizations. *Review of Philosophy and Psychology, 7*(1), 163–180. doi:10.1007/s13164-015-0264-1

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences, 108*(17), 6889–6892. doi:10.1073/pnas.1018033108

Diesen, C. (2001). The advantages and disadvantages of lay judges from a Swedish perspective. *Revue Internationale De Droit Pénal, 72*(1), 355–363. doi:10.3917/ridp.721.0355

Fehr, B. (1988). Prototype analysis of the concepts of love and commitment. *Journal of Personality and Social Psychology, 55*(4), 557–579. doi:10.1037/0022-3514.55.4.557

Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology, 66*(2), 127–148. doi:10.1037/0021-9010.66.2.127

Goldbach, T. S., & Hans, V. P. (2014). Juries, lay judges, and trials. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp.2716–2727). New York, NY: Springer Science and Business Media.

Goodman-Delahunty, J., Granhag, P. A., Hartwig, M., & Loftus, E. F. (2010). Insightful or wishful: Lawyers’ ability to predict case outcomes. *Psychology, Public Policy, and Law, 16*(2), 133–157. doi:10.1037/a0019060

Government of Sweden (1998 [1942]). The Swedish Code of Judicial Procedure (Ds 1998:65). https://www.regeringen.se/rattsliga-dokument/departementsserien-och-promemorior/1998/01/ds-199865/

Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2007). Blinking on the bench: How judges decide cases. *Cornell Law Review, 93*(1), 1–44.

Horowitz, L. M., & Turan, B. (2008). Prototypes and personal templates: Collective wisdom and individual differences. *Psychological Review, 115*(4), 1054–1068. doi:10.1037/a0013122

Hovland, C. I., & Sherif, M. (1980). *Social judgment* (Reprint from 1961 ed.). Greenwood, Westport, Conn.: Greenwood.

Inman, M. L., & Baron, R. S. (1996). Influence of prototypes on perceptions of prejudice. *Journal of Personality and Social Psychology, 70*(4), 727–739. doi:10.1037/0022-3514.70.4.727

Kalai, E. (1993). A rational game theory framework for the analysis of legal and criminal decision making. In R. Hastie (Ed.), *Inside the juror: The psychology of juror decision making* (Vol. Cambridge Series on Judgment and Decision Making, pp. 235–241). New York, NY: Cambridge University Press.

Krumm, A. J., & Corning, A. F. (2008). Perceived control as a moderator of the prototype effect in the perception of discrimination. *Journal of Applied Social Psychology, 38*(5), 1109–1126. doi:10.1111/j.1559-1816.2008.00341.x
Lakoff, G. (1987). *Women, fire and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago Press.

Lindholm, T. (2008). Who can judge the accuracy of eyewitness statements? A comparison of professionals and lay-persons. *Applied Cognitive Psychology, 22*(9), 1301–1314. doi:10.1002/acp.1439

Martz, E., Strohmer, D., Fitzgerald, D., Daniel, S., & Arm, J. (2009). Disability prototypes in the United States and the Russian Federation: An international comparison. *Rehabilitation Counseling Bulletin, 53*(1), 16–26. doi:10.1177/0034355208329357

McCaughey, T. J., & Strohmer, D. C. (2005). Prototypes as an indirect measure of attitudes toward disability groups. *Rehabilitation Counseling Bulletin, 48*(2), 89–99. doi:10.1177/00343552050480020301

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*, 89–115. doi:10.1146/annurev.ps.32.020181.000513

Nathan, B. R., & Lord, R. G. (1983). Cognitive categorization and dimensional schemata: A process approach to the study of halo in performance ratings. *Journal of Applied Psychology, 68*(1), 102–114. doi:10.1037/0021-9010.68.1.102

Pettersson, M., Dahlman, C., & Sarwar, F. (2018). År det sant att lekmannadomare påverkas av juridiskt irrelevanta omständigheter, men inte juristdomare? *Juridisk Tidskrift Vid Stockholms Universitet, 2017-18*(2), 363–376.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*(3), 192–233. doi:10.1037/0096-3445.104.3.192

Tetlock, P. E. (1983). Accountability and the perseverance of first impressions. *Social Psychology Quarterly, 46*(4), 285–292. doi:10.2307/3033716

Toye, R. (2013). *Rhetoric: A very short introduction*. Oxford, UK: Oxford University Press.

Winter, S. L. (1988). The metaphor of standing and the problem of self-governance. *Stanford Law Review, 40*(6), 1371–1516. doi:10.2307/1228780

Winter, S. (2001). A clearing in the forest—Law, life and mind. Chicago: University of Chicago Press.

Wright, R., Junious, T., Neal, C., Avello, A., Graham, C., Herrmann, L., & Walton, N. (2007). Mental fatigue influence on effort-related cardiovascular response: Difficulty effects and extension across cognitive performance domains. *Motivation and Emotion, 31*(3), 219–231. doi:10.1007/s11031-007-9066-9