

1 Visual Vignettes for Cross-National Research

2
3 Matthew Facciani*

4 Matthew E. Brashears

5 Jingwen Zhong

6 University of South Carolina

7 Word count: 8,099

8 Keywords: vignettes, visual, methodology, cross-national

9
10
11
12
13
14
15
16
17
18

* Direct correspondence to: Matthew Facciani, Department of Sociology, University of South Carolina, Sloan College. Rm. 321, 911 Pickens St. Columbia, SC 29208. facciani@email.sc.edu. We would like to thank Dr. Andrea Henderson and Dr. Eric Gladstone for their helpful comments on our manuscript. This research was supported by a grant from the Army Research Institute (#W911NF-13-R-0001).

19 **Abstract:**

20 Textual vignettes are a widely popular research tool, but they assume literacy and reading
21 comprehension skills that limit their utility for both research with disadvantaged populations and
22 cross-national research.

23 We describe a new visual vignette approach, based on approaches used in studying populations
24 with lower verbal ability, that avoids the issues raised by live actors while also reducing reliance
25 on written language typical to textual vignettes.

26 We examine the effectiveness of our new visual vignette using a survey experiment in which
27 respondents were exposed to either a visual vignette, or a textual vignette, depicting the same
28 interaction. We found that our visual vignette produced similar results when compared to
29 traditional textual vignettes, confirming their utility. Additionally, we report on a pilot data
30 collection using a cross-national sample, in which English literacy cannot be assumed. Our
31 results indicate that our new method should improve research options both with low literacy
32 populations and foreign language speakers.

33 **Introduction**

34 How can we use vignettes to collect data from persons who speak different languages or
35 have low literacy? Researchers often want to collect data from disadvantaged populations or to
36 extend research beyond the classic Western, educated, industrialized, rich and democratic
37 (WEIRD) populations used in many experiments (Henrich, Heine, and Norenzayan, 2010). Yet,
38 doing so requires grappling with imperfect literacy, and in the case of cross-national research,
39 can involve expensive translations that may or may not convey identical subtexts across all
40 cultures. Building on research about autistic and preverbal populations, we generate a new
41 “visual” vignette approach in order to solve these problems. By presenting interactions visually,
42 we avoid the problems presented by imperfect literacy or translation. Our approach should
43 facilitate the collection of standardized data from both low literacy respondents, as well as cross-
44 national research for participants who speak different languages.

45 We develop a textual vignette that describes a simple social situation involving a variety
46 of status cues during a work meeting. We then created an animated version of this vignette that
47 portrays the same behaviors in visual form. By comparing the results of a visual and textual

48 vignette approach, we are able to validate the effectiveness of our visual vignette approach. If
49 our visual vignette method is successful, the characters in each vignette type should be rated
50 similarly by participants. We also describe a pilot data collection using a cross-national sample
51 as a proof-of-concept for the appropriateness of our method. We describe our new visual vignette
52 approach, introduce our data, present our results, and conclude with directions for future
53 research.

54 **Background**

55 Vignettes, or short paragraphs of text describing a set of characters and/or a short
56 sequence of events, are a widely used research tool in psychology (e.g. Barrera & Buskens,
57 2002), sociology (e.g. Alves & Rossi, 1978), marketing (Wason, Polonsky, & Hyman, 2002),
58 and education (e.g. Anast Seguin & Ambrosio, 2002) research. Atzmüller and Steiner (2010, p.
59 128) define vignettes as “a short, carefully constructed description of a person, object, or
60 situation, representing a systematic combination of characteristics.” For example, the vignette
61 could be a written story about several characters interacting with each other, and the participant
62 must rate how appropriate each character’s behavior was in the scenario. Typically, researchers
63 construct the vignette so as to make a factor of research interest relevant, and present a variety of
64 survey items after the vignette to determine how participants evaluate and react to the vignette
65 content. Using vignettes in this fashion is formally known as experimental vignette methodology
66 (EVM), and is most suitable for establishing control over independent variables and improving
67 causal inference (Aguinis & Bradley, 2014). EVM allows researchers to present realistic
68 scenarios to their participants while manipulating their variables of interest in an experimental
69 setting.

70 One common type of EVM involves presenting vignettes to participants and then asking
71 them to answer questions about the characters depicted (Aguinis & Bradley, 2014). Most often
72 vignettes in these studies are presented as blocks of text, which requires participants to be literate
73 (Hughes, 1998). But this necessarily imposes limitations when studying certain populations, such
74 as the reading impaired or those who speak a different language than the researchers. For the
75 former, researchers could resort to auditory presentation of vignette information. But for the
76 latter, it is necessary to translate vignettes into a different language. Unfortunately, differences in
77 meaning or connotation between languages can make this a very difficult, and expensive,
78 process. Video vignettes, or vignettes created using full motion video of human confederates, are
79 a useful alternative to reach these populations.

80 Video vignettes, or recorded depictions of living individuals acting out scripted events,
81 can reduce the reliance on participant literacy and even add authenticity to the situation (Simon
82 and Tierney, 2011). Video vignettes were found to help supplement text vignettes and improve
83 learning outcomes in medical residents while studying patient cases (Balslev et al, 2005).
84 Additionally, medical students found videos of patient cases more memorable than purely text-
85 based vignettes (De Leng et al, 2007). Sled and colleagues (2002) examined how often
86 participants would blame a female character for her own sexual assault, finding that the victim
87 was blamed much more often by subjects exposed to a textual vignette than by subjects exposed
88 to a video vignette. The researchers argued that the textual vignette condition may have
89 facilitated the use of existing myths about rape, and therefore led to greater blame for the victim,
90 while the video vignettes permitted less ambiguity about the situation and as a result provided
91 the experimenter with more control over participant perceptions of the situation. Thus, evidence
92 suggests that video vignettes produce more salient outcomes than purely text-based vignettes.

93 While video vignettes offer certain advantages compared to textual vignettes, they come
94 with their own limitations. Firstly, presenting information in a video recording has been shown to
95 add noise (i.e. background information that is irrelevant to the main point of the story) to the
96 manipulation, which can decrease recall accuracy (Woehr & Lance, 1991). Additionally, it is
97 expensive to hire actors and compose an authentic portrayal of a situation. Even if the researcher
98 has enough funding to pay actors, an effective video vignette requires believable settings and
99 good quality photography, or else subjects may be unable to experience the video as if it were a
100 real interaction. As Tippett (2018) found when examining the effectiveness of sexual harassment
101 training, if videos are seen as inauthentic or poorly acted, it can distract viewers from the
102 content. If the video is experienced as inauthentic, then it is hard to argue that the interactions
103 depicted will impact the respondent as the researcher intends. Even if this issue can be avoided,
104 the characteristics of the actors can introduce confounding effects in the study. Real actors come
105 as collections of traits, including height, weight, skin tone, attractiveness and accent, among
106 others, which provide a host of cues to subjects about the probable background of the “character”
107 (Heir 2000; Marlowe, Schneider & Nelson, 1999). In contrast to textual vignettes, these traits
108 cannot simply be omitted in a video vignette and so will be present, and potentially influential.
109 Unless these traits are very carefully controlled, and the research deriving from the “thin slicing”
110 paradigm implies that this is a tall order, confounds are likely to enter even the best planned
111 study (Ambady, Bernieri & Richeson, 2000). Efforts to replace human actors with realistic
112 computer-generated characters (e.g., O’Connor and Gladstone, 2015) are helpful, but do not
113 eliminate this problem. While textual vignettes have the option of simply not mentioning
114 particular details (e.g., skin color), these same details cannot be omitted from realistic

115 animations, and attempts to do so run the risk of inducing an “uncanny valley” reaction (e.g.,
116 Mathur and Reichling, 2016; Steckenfinger and Ghazanfar, 2009).

117 To avoid the language challenges inherent to textual vignettes, and the expense and
118 confounds of using real actors in video vignettes, we propose a middle path. Interactions can be
119 depicted in visual format using a set of characters that are deliberately non-human in appearance,
120 but are still experienced as social actors. Studies have shown that humans attribute agency and
121 intention to non-living objects (Castelli et al, 2000; Kelemen, 1999). The Social Brain
122 Hypothesis (Dunbar, 1998; Whiten & Byrne, 1997) theorizes that humans’ brains have evolved
123 to solve social, rather than physical problems. Thus, neurotypical¹ humans should have a
124 proclivity for experiencing a wide range of environmental stimuli in social terms. Put differently,
125 if the selection pressures leading to high general problem solving intelligence in humans stem
126 primarily from the need to reason about social situations, then we ought to observe a tendency to
127 process social content more efficiently than non-social (e.g., Janicik and Larrick, 2005; Simpson,
128 Markovsky and Steketee, 2011), and to tackle non-social problems in a social manner. Consistent
129 with this work, we rely for inspiration on a seminal paper by Heider and Simmel (1944), which
130 found that moving shapes will often be spontaneously interpreted in social terms. Their simple
131 task presented a short film of shapes moving around on a screen and asked respondents to
132 interpret what was seen. Neurotypical individuals were more likely to describe the events in
133 explicitly social terms, even though the social “actors” in these videos were simple geometric
134 shapes. This task has subsequently been used to study the human ability to understand the mental
135 states of others (i.e., theory of mind), and has been adapted as an assessment for autism spectrum

¹ This term typically refers to those without a clinically diagnosed mental illness or neurological disorder. For our purposes, we are focusing on those who are not on the autism spectrum.

136 disorder (e.g., Baron-Cohen 2000; Klin, 2000). Using these results, we argue that a visual
137 vignette depicting interactions between minimally humanized geometric shapes will both avoid
138 the literacy requirements of textual vignettes, as well as the expense and confounds of video
139 vignettes.

140 In short, because human cognition is biased to understand events in social terms, it is
141 possible to present social interactions using clearly non-human actors without compromising
142 validity. The result is an easy to interpret stimulus item that can be used with a variety of
143 populations without the need for extensive translation and back-translation. Our visual vignette
144 method could be useful for studies involving non-English speaking, but literate populations via
145 the internet as well as non-literate populations via interviews with tablets. Moreover, as these
146 simple visual vignettes impose much lower computing requirements than full motion video, they
147 can be used with tablet-administered surveys and data collections occurring as part of field
148 interviews. As such, while they can be administered via the internet, they can also be used in
149 offline contexts. Additionally, once the main visual vignette has been created, the researchers can
150 apply minor edits with the characters, situation, and setting to implement a variety of conditions.
151 The current study tests this novel visual vignette approach and compares its effectiveness to a
152 more traditional textual vignette describing the same interactions.

153 **Current Study**

154 We developed visual and textual vignettes and assessed whether participants judged the
155 depicted scenarios similarly. We adapt our vignette from one used by Fisek and Hysom (2008)
156 who employed vignettes to assess status characteristics and reward expectations in a work
157 setting. Our purpose is not specifically to assess status characteristics theory, but this vignette
158 provides a previously tested stimulus to make our methodological assessment more reliable.

159 Status cues are “indicators, markers, or identifiers of the different social statuses people possess”
160 (Berger et al 2018, p.155). These status cues can be symbolic items possessed by individuals,
161 behaviors that convey status and enhance perceived competence, or both. These behavioral
162 signals are often widely understood by members of the same society and so should provoke
163 consistent reactions. While different cultures often possess unique status cues, essentially all
164 human cultures incorporate a concept of status and status signaling. Therefore, it should be
165 possible to measure the extent to which different cultures share specific cues via our method. A
166 vignette showcasing various status behaviors, followed by an instrument measuring the
167 perceived status of the characters, would allow researchers to determine if the same behaviors
168 convey similar or even identical perceptions of status across different cultures.

169 The social situation for our vignette incorporates two well-studied status cues in the
170 literature: choosing to sit at the head of a table and interrupting others. Choosing to sit at the
171 head of the table has been found to signal status (Hare & Bales, 1963; Lott & Sommer, 1967),
172 which increases the perceived competence of those performing this behavior (Nemeth &
173 Wachtler, 1974). Interrupting is viewed as an inappropriate social behavior with those who
174 interrupt being rated as disrespectful, rude, and confrontational (Farley 2008; LaFrance 1992).
175 Both of these behaviors are well-studied in previous research and are easy to implement in both
176 textual and visual vignettes.

177 Yadav and colleagues (2011) found that stories presented in video formats resulted in
178 greater emotional engagement by participants compared to stories in text only formats.
179 Furthermore, Slead and colleagues (2002) found that video vignettes created less ambiguity
180 about the social situation as they allowed the participant greater control over the focus of their
181 attention. Thus, we think it is possible that our visual vignette may produce stronger effects in

182 the character ratings compared to the textual vignette. We propose the following competing
183 hypotheses:

184 *Visual Similarity Hypothesis:* the textual and visual vignettes should produce similar effects for
185 the judgements of character behaviors

186 *Visual Strength Hypothesis:* the visual vignette should produce even stronger effects for the
187 judgement of character behaviors when contrasted against the textual vignette

188 We are also interested in possible gender effects. While interrupting has been shown to
189 be an inappropriate behavior, it is often regarded as much more appropriate for a man to interrupt
190 than a woman (Aries 1987; Rosenblum, 1986). Farley (2008) demonstrated that women who
191 interrupt are rated as less likable than men who engage in the same interrupting behaviors. Thus,
192 to examine possible gender effects, we produced two versions of each textual and visual vignette,
193 which were identical except for exchanging the names of specific characters who engaged in the
194 specified behaviors. We are therefore able to hold the behaviors constant, while varying the
195 perceived genders of the characters performing those behaviors.

196 In the first vignette, a female character sat at the head of the table, another female
197 character was ignored and interrupted, a male character did the ignoring, and finally another
198 male character did the interrupting. In the second vignette, a male character sat at the head of the
199 table, another male character was ignored and interrupted, a female character did the ignoring,
200 and finally another female character did the interrupting (see Appendix A).

201 Once we constructed the textual vignette, we hired an animator to translate the same
202 vignette into visual form. In addition to the geometric figures derived from the Heider-Simmel
203 task (1944), we also included simple facial expressions similar to those used by Hamlin and

204 Bloom (2007) to evaluate the social awareness of nonverbal infants. These facial expressions,
205 relying solely on simple eyes and mouth shapes, convey significant meaning while being generic
206 enough to mitigate any attractiveness biases. Our visual vignette characters were all colored
207 green to avoid any extant human skin tone. While various colors may provoke particular
208 reactions, none of these reactions are likely to be social (given that humans are not green), and in
209 any event character color is constant across all conditions. We also added speech bubbles,
210 containing text-like scribbles, to convey speech and a generic speech-like sound (i.e., gibberish
211 with the approximate cadence of language) while the mouths were moving. This speech-like
212 sound takes advantage of the audio capabilities of many tablets, computers and smartphones, but
213 is not needed to convey speech. Like the textual vignette, our visual vignette was produced in
214 two versions, differing only in that the names of the characters were swapped such that men and
215 women were engaging in opposite behavior. Names were changed in the textual vignette and
216 nametags were changed in the visual vignette. Importantly, we used common American names
217 because our initial sample was drawn from the United States, and because US global media
218 dominance maximizes the chances that respondents from any country would recognize the sex
219 typically associated with particular names. Obviously, the names could be adjusted to match the
220 local context, but this would require changes to the animation itself. A single still frame from our
221 visual vignette, showing all characters as well as the simple background, is included in Figure 1
222 (see supplemental material for full visual vignettes).

223 ---insert Figure 1 here---

224 Finding support for our *Visual Similarity Hypothesis* would suggest that our new visual
225 vignette is working as intended and is a candidate vignette type for cross-national research. If the
226 visual vignette elicits stronger reactions when compared to the text vignette, then our *Visual*

227 *Strength Hypothesis* would be supported. Finally, we are also testing if the same gender effects
228 are seen in both types of vignettes, which would also support our *Visual Similarity Hypothesis*.
229 Therefore, our hypotheses, and corresponding empirical tests, provide a comprehensive
230 assessment of the efficacy of our new approach².

231 **Methods**

232 *Subjects* – We employ two distinct samples to validate our approach. Our first sample
233 draws participants from the United States only while our second sample includes additional non-
234 US countries. If our visual vignette produces similar or stronger reactions as traditional textual
235 vignettes with a US population then the validity of this approach will receive support. If our
236 visual vignette is then able to generate sensible results with our non-US sample, then it suggests
237 that it will be efficacious with low literacy and non-English speaking groups. In all cases,
238 participants were limited to those who were minimally literate in English (i.e., literate enough to
239 participate in subject recruitment, described below). This is a limitation of our data, imposed by
240 difficulties in obtaining more representative data with our available budget. However, we note
241 that it is much easier to translate (or understand in a second language) a direct question than an
242 entire textual vignette. As such, even if the questions continue to be posed in English, we argue
243 that presenting the vignette in visual form will improve validity.

244 For the first sample, participants were recruited from Amazon’s Mechanical Turk online
245 population. Mechanical Turk is an online platform where participants seek out tasks to complete
246 for compensation. Participants create online accounts and choose from a variety of tasks that are
247 simple for humans but difficult or impossible for automated systems (e.g., classifying photos).

² We did not create a control condition without status cues because we felt it would be too different of a social situation to make adequate comparisons. However, between conditions status effects should be sufficient to indicate the operation of status cues.

248 Multiple research teams have found this to be an effective way to recruit subjects for research
249 tasks, exchanging some degree of experimental control for a more diverse subject pool than is
250 usually available on campus (Buhrmester, Talafar, & Gosling, 2018). Researchers can restrict
251 participation to certain criteria including global place of residence. Previous research finds that
252 Mechanical Turk workers produce data of a quality that is comparable to recruiting via social
253 media sites and in-person populations (Casler, Bickel, & Hackett, 2013), and this is especially
254 true of high-reputation workers, or workers who have been assessed by previous task-offerers as
255 being especially devoted (Peer et al, 2014). Mechanical Turk workers are also more
256 socioeconomically and ethnically diverse compared to the college student samples frequently
257 used in experimental research (Casler, Bickel, & Hackett, 2013) and therefore enhance
258 generalizability. Participants were paid \$1 USD to participate in the task and all tasks took
259 approximately ten minutes to complete. This pay rate is competitive with other Mechanical Turk
260 studies (Hara et al, 2018). No subject was allowed to participate in more than one version of the
261 task. We restricted our Mechanical Turk sample to those who lived in the United States and
262 completed tasks at a performance of 90% or greater to obtain high-reputation workers in our
263 sample. Our Mechanical Turk participants all spoke English and were generally well-educated
264 (55% had college degree or higher, 59% were male, mean age: 33.68). While our task is
265 intended to facilitate cross-national research, initially validating its usefulness with a uniform,
266 relatively well understood population is more straightforward than the alternatives.

267 *Procedure* – Participants were assigned to one of two versions of either the text or visual
268 vignettes (creating four total conditions). Version 1 (for both text and visual vignettes) had the
269 female character sit at the head of the table and the male character interrupt. Version 2 (for both

270 text and visual vignettes) had the male character sit at the head of the table and the female
271 character interrupt (see online Supplementary Table 1).

272 We included scales of appropriateness and expectations of competence for our
273 participants to rate the vignette characters. Status implies competence and thus this provides an
274 indirect measure of status effects. Likewise, behaviors such as interrupting are frequently faux
275 pas, and so will be judged as inappropriate by observers. However, the degree of
276 inappropriateness should vary by the characteristics of the interrupter and interrupted parties.
277 The Likert scales for both appropriateness and competence ranged from totally inappropriate to
278 totally appropriate on a 1 to 7 scale. The Likert scale for expectations of competence was a 4-
279 item index ($\alpha = 0.82$) used by (Thye & Harrell, 2017) with higher competence levels suggesting
280 greater status through a higher level of task ability. The Likert scale measuring appropriateness
281 was generated for this study to assess how respectful and civil an individual was in a situation,
282 since interrupting has been shown to result in perpetrators being rated as disrespectful, rude, and
283 confrontational (Farley, 2008; LaFrance, 1992). This measure was also inspired by the concept
284 of *deflection*, from Affect Control Theory (Heise, 2007). Deflection is a sense of unreality or
285 inappropriateness that individuals experience in social situations and is predictable based on a set
286 of established equations and measurements. Conceptually, deflection measures how much
287 distance there is between what an individual expects to happen in a social situation from cultural
288 sentiments and what they actually experience. Behaviors that are rated as appropriate ought to be
289 low-deflecting, whereas inappropriate behaviors are higher-deflecting (i.e., a certain type of actor
290 is behaving in a way that they should not).

291 For the textual vignette conditions, participants first read the vignette and then rated the
292 characters' appropriateness and general competence. For the visual vignette conditions,

293 participants were told they would watch a short video of a social situation and would then be
294 asked about the characters in the story. The participants had to watch the video fully at least
295 once, but were allowed to replay the video as much as they wanted to.³ While answering
296 questions, subjects were able to observe a still image from the vignette (see Figure 1) to act as a
297 memory aid. Participants in the visual vignette conditions received the same questions about
298 appropriateness and competence as respondents in the textual vignette versions. After rating the
299 characters, participants then answered a few demographic questions and were debriefed.

300 We used a series of ANOVAs with Bonferroni post-hoc tests to evaluate if the character
301 who sat at the head of the table was judged as more competent than the other characters for both
302 vignette types. We additionally evaluated whether the character who interrupted was rated as less
303 appropriate than the other characters in both vignette types. We also compared the means of
304 character ratings for each vignette type to determine if vignette type created a difference in
305 character ratings. Finally, we compared means via t-tests of character ratings from vignette
306 version 1 and 2 in each vignette type to identify any gender effects.

307 **Results**

308 *Textual Vignettes*

309 In version 1 of our textual vignette, the character who interrupted was rated as less
310 appropriate than the character who was sitting at the head of the table and the character who was
311 interrupted ($F = (3, 236) 14.90; p < .01$). The negative impact of interrupting on character
312 appropriateness was also significant compared to the character who was not looking ($p < .05$;
313 one-tailed). Additionally, the character who sat at the head of the table was rated as more

³ There was a maximum of 15 replays allowed, but most participants viewed it 1-2 times and the maximum amount viewed was 3 times.

314 competent than the characters who interrupted and ignored the others ($F = (3, 236) 10.25; p$
315 $<.01$).

316 Version 2 of the textual vignette swapped the character's genders and our analysis
317 revealed that the female character who interrupted was still rated as less appropriate than the
318 character who sat at the head of the table ($F = (3, 236) 5.40; p <.01$). The difference between the
319 female character who interrupted and the character who was being interrupted was also
320 significant ($p <.01$). Finally, there were no significant differences for competence between the
321 characters ($F = (3, 236) 1.53; p <.05$). Overall, these results are consistent with expectations
322 based on prior research and indicate that our textual vignettes are behaving appropriately (see
323 online Supplementary Tables 2-5 for full details). Next, we check if our visual vignettes produce
324 similar results.

325 *Visual Vignettes*

326 Visual vignette version 1 revealed that the character who interrupted was rated as less
327 appropriate compared to the character who sat at the head of the table as well as the character
328 who was being interrupted ($F = (3, 236) 10.84; p <.05$). Additionally, the character who sat at the
329 head of the table was rated as more competent than the interrupter ($F = (3, 236) 13.26; p <.01$)
330 and the character who was not looking ($p <.05$; one-tailed).

331 For visual vignette version 2, we find that the female character who interrupted was rated
332 as less appropriate than the character who sat at the head of the table ($F = (3, 236) 22.67; p <.01$).
333 We also found that version 2 of the visual vignette revealed that the character who sat at the head
334 of the table was rated more competent than the other character who was interrupting ($F = (3,$
335 $236) 9.36; p <.01$) and the character who was not looking ($p <.05$; one-tailed). Thus, both the

336 textual and the visual vignette generally found that characters sitting at the head of the table are
337 rated with greater appropriateness and characters interrupting others are rated with reduced
338 appropriateness (see online Supplementary Tables 6-9 for full details). These results support our
339 *Visual Similarity Hypothesis* and indicate that our respondents react as anticipated to
340 conventional textual vignettes, and that they exhibit similar responses to our new visual
341 vignettes, and therefore supports the usefulness of our new approach.

342 *Textual vs. Visual Vignettes*

343 We compared the means of character ratings between the textual and visual vignettes via
344 a series of t-tests (see online Supplementary Table 10 for full details). In the first version, we
345 found that the character who was interrupting was rated less competent ($t = -1.74$; $p < .05$; one
346 tailed), but not less appropriate in the visual vignette. However, the male character who was not
347 looking and ignoring the female character was not rated differently in either vignette type.
348 Additionally, the character being interrupted was rated more competent ($t = 1.84$; $p < .05$, one
349 tailed) and more appropriate ($t = 2.38$; $p < .05$) in the visual vignette. However, the character
350 who sat at the head of the table was not rated as more appropriate or more competent in the
351 visual vignette.

352 When we compared the version 2 text and visual vignettes we found several significant
353 differences. Unlike in version 1, the character who was sitting at the head of the table was rated
354 as more appropriate ($t = 2.65$; $p < .05$), but not more competent in the visual vignette. This may
355 reflect that in the textual vignette readers are able to infer additional behavior (e.g., posture)
356 whereas in the visual vignette researchers can control these variables more precisely.
357 Furthermore, the character who was interrupting was rated as less competent ($t = -2.78$; $p < .05$)
358 in the visual vignette, but not less appropriate much as in version 1. Additionally, unlike in

359 version 1, the female character who was not looking at and ignoring the male character was rated
360 less competent ($t = -2.09$; $p < .05$). Finally, the character who was interrupted was again rated
361 more appropriate in the visual vignette ($t = 2.11$; $p < .05$), but not more competent. Our visual
362 vignette therefore produces stronger appropriateness for sitting at the head of the table (in
363 Version 2 with the male character) as well as reduced competence for interrupting other
364 characters (in both versions with male and female characters) partially supporting our *Visual*
365 *Strength Hypothesis*. This is especially noteworthy given that, as pointed out in footnote 3, most
366 respondents watched the video 1-2 times, and so we produced similar or stronger effects as
367 textual vignettes even from a very small amount of exposure to a visual vignette. The
368 interruption behavior appeared more salient in the visual vignette. Moreover, the greater strength
369 of the visual vignette stimulus should serve to allow effective studies using this methodology
370 even with somewhat smaller samples sizes than is typical.

371 Lastly, we used t-tests to contrast the results of versions 1 and 2 to assess the gender
372 effects of the characters for both textual and visual vignettes. Unfortunately, we did not find any
373 significant results except for version 1 of the textual vignette, where the female character was
374 rated as more appropriate than the male character when sitting at the head of the table ($t = 1.85$; p
375 $< .05$; one-tailed). There were no other significant gender effects for the textual or visual
376 vignettes. Thus, neither vignette style appears to be significantly impacted by gender effects.

377 In total our data lead to a clear conclusion: our visual vignettes are able to produce the
378 same reactions in subjects, and in some cases stronger reactions, despite the lack of photorealistic
379 (or even particularly accurate) representations of human beings. This is an important validation
380 of our method and suggests that it is a viable means of collecting vignette style data cross-
381 nationally or with literacy challenged populations.

382 *Multicultural data*

383 Our Mechanical Turk data were drawn only from participants within the United States so
384 that we could confirm that the visual vignette strategy was workable with a known population.
385 But a core goal of our effort is to see if our visual vignette effects hold for participants from a
386 variety of cultural backgrounds. Difallah, Filatova, and Ipeirotis (2018) found that over 90% of
387 MTurk workers came from just two countries (United States: 75%; India: 16%). Thus, for our
388 second sample it was necessary to use a platform that can recruit participants from a more
389 diverse assortment of countries. Therefore, as a proof of concept, we also collected data using the
390 Volunteer Science⁴ platform. Volunteer Science is a science crowdsourcing project that recruits
391 individuals through various social media platforms to participate in scientific research without
392 payment. It is thus similar to other scientific crowdsourcing platforms, such as the Zooniverse
393 Project⁵, and provides greater geographic reach, and fewer concerns about confounds from profit
394 motivation, relative to MTurk. Our data collection yielded participants from 54 different
395 countries (see supplemental material), with no more than six participants from any one country.
396 These participants all spoke English and were generally well-educated (40% had college degree
397 or higher, 55% were female, mean age: 27.97). Our Volunteer Science participants were less
398 educated, were more likely to be female, and were slightly younger than our Mechanical Turk
399 sample. Our data are thus geographically diverse, but we lack the necessary per country sample
400 size for legitimate statistical comparisons across different regions, and our respondents are
401 almost certainly not representative. The Volunteer Science participants were all proficient in
402 English so this will not test whether our vignette can be used with low literacy populations or

⁴ <https://volunteerscience.com>

⁵ <https://www.zooniverse.org>

403 with other languages. However, these more diverse data should still allow us to test the
404 plausibility of using visual vignettes to collect data from non-US, and non-Western, populations.
405 And, as noted above, it is easier to read and understand direct questions in a second language,
406 than to read and grasp the nuances in a textual vignette. The participants completed only the
407 visual vignette task (version 1) with an identical procedure as the Mechanical Turk sample.

408 In our Volunteer Science data, we found that the female character who was being
409 interrupted was rated higher in competence compared to the character who was ignoring ($F = (3,$
410 $396) 3.05; p < .05$). However, there were no significant differences in competence for the
411 character sitting at the head of the table or interrupting. When looking at appropriateness ratings,
412 we did find that the female character who sat at the head of the table was rated with greater
413 appropriateness than the other three characters ($F = (3, 396) 17.84; p < .001$). Additionally, the
414 male character who interrupted was rated with lower appropriateness compared to the character
415 at the head of the table and the character who was being interrupted ($F = (3, 396) 17.84; p$
416 $< .001$). Thus, our multicultural sample produced anticipated results for our characters in terms of
417 appropriateness, but not competence (see online Supplementary Tables 11-12 for full details).
418 This may indicate greater variation in the behaviors that signal competence across nations, but
419 the small number of persons per country in our sample makes it impossible for us to test this
420 possibility. Our multicultural data provides evidence that is generally supportive of our
421 multicultural hypothesis, though more extensive cross-national testing needs to be done.

422 **Discussion**

423 Can a visual vignette using shapes with minimal humanizing features elicit similar effects
424 when compared to more traditional textual vignettes? Our experiments indicate that the answer is
425 yes, as both our textual and visual vignette produced the anticipated status effects from the

426 characters. The character who sat at the head of the table was rated as more appropriate and
427 competent in both the visual and textual vignettes, while the character who interrupted another
428 character was rated as less appropriate supporting our hypotheses for predicted status effects.

429 Not only did our visual vignette produce the expected results, it may have produced more
430 potent effects as well. The male character sitting at the head of the table was rated with greater
431 appropriateness in the visual vignette for Version 2 only. Additionally, the character who was
432 interrupted was rated as more appropriate in the visual vignette of both versions (and more
433 competent in version 1). Perhaps actually seeing the interruption take place increases sympathy
434 as there is less room for ambiguity. In version 2, the female character who was not looking at the
435 other character, as well as the female character who was interrupting the other, were both rated
436 as less competent. Again, actually seeing this behavior could produce a more substantial effect.
437 There may also be a possible gender influence, as the male characters who ignored and
438 interrupted were not rated differently between textual and visual vignettes. In other words, it may
439 be that ignoring and interrupting others is a sufficiently standard “male” behavior that, to a
440 degree, can evade negative reaction from observers.

441 Our multicultural data also found that watching a visual vignette where a character is
442 sitting at the head of the table increases that character’s perceived appropriateness while
443 watching a character interrupt another decreases their appropriateness. Thus, our visual vignette
444 shows promise for working across more diverse samples. Despite these positive results, we did
445 not find a gender effect that female characters were not rated less appropriate than male
446 characters when they interrupted.

447 While our study offers a promising tool for future research, it is not without several
448 limitations. First, our study was limited to Mechanical Turk workers in the United States and

449 participants in the Volunteer Science program. All these participants spoke English and still had
450 to answer questions about the characters with some degree of textual proficiency. Future work
451 should be done with non-English literate populations, but we note that in testing our approach
452 with even this limited multi-cultural sample, we have diversified the respondents considerably
453 relative to most comparable efforts. It's important to note that our method does not have to be
454 administered via the internet, but could also be done via tablets in the field. It therefore has
455 promise to aid field research with a variety of populations, and especially with non-literate
456 populations who can respond to questions verbally, while still receiving an experimentally
457 precise manipulation.

458 Additionally, the gender variation between our vignette versions was accomplished
459 through name changes, which may have been too gentle a manipulation. A more salient gender
460 manipulation could be obtained by calling more attention to the gender of the characters in the
461 vignette, for example by adding secondary sexual characteristics to the visual vignette. As a
462 consequence, we failed to observe the gender effects that we anticipated based on prior research.
463 While this is inconsistent with previous work (Aries 1987), it's possible that our Mechanical
464 Turk population being highly educated and fairly young reduced the likelihood that they
465 possessed antiquated gender stereotypes. Although some studies show that merely changing the
466 name of a character can produce a gender effect (Bongiorno, Bain & David, 2014), Heflick and
467 colleagues (2011) found that focusing on the physical appearance of men and women can
468 produce gender effects, with men being viewed as more competent than women. Thus, the lack
469 of obvious secondary sexual characteristics or gender typical grooming or fashion in our visual
470 vignettes might have prevented us from observing gender effects. Additionally, our American
471 names may have created a confounding variable for our multicultural sample. Due to this sample

472 being from many different countries, it was not possible to pick names that were most common
473 in each. If researchers aim to apply this method to specific cultures, then they should name the
474 characters with common names from those cultures. Yet, by using common American names, we
475 maximized our chance that US media dominance would allow all participants to correctly infer
476 the intended sexes of the characters. Finally, the small number of viewings of the visual vignettes
477 by most participants, while sufficient to produce overall status effects, may not have been
478 enough to completely activate gender expectations.

479 Our new visual vignette technique is most suited for situations that can be depicted
480 visually. For example, our visual vignette can show characters moving around in a physical
481 setting, but traditional textual vignettes may have an advantage in describing the mental states of
482 characters (e.g., Cindy was worried about her upcoming presentation). It should be kept in mind,
483 however, that in most real-world interactions individuals are unable to directly access the
484 thoughts and feelings of interactants, instead relying on context, body language, and other visible
485 indicators. As such, we think this is a challenge that should be addressable with additional work,
486 rather than a fundamental limitation to the method. Indeed, by forcing researchers to focus on
487 what respondents in real interactions *can* perceive, our method may improve the generalizability
488 of the results.

489 Our visual vignette approach possesses useful economic advantages. Whereas
490 photorealistic (e.g. motion capture) animation techniques require substantial equipment and
491 expense, undergraduates studying graphic design will frequently have the skills required to
492 generate basic visual vignettes. Likewise, universities may own a site license for common
493 animation software or make it available at reduced cost (e.g., our animator used Adobe
494 Animate). While generating visual vignettes to researcher specification may require some

495 research funds, the cost of generating these visuals is far less than more sophisticated alternatives
496 (e.g., we paid our student animator \$200 for our visual vignettes). The primary limitation is thus
497 the creativity of researchers in finding ways to present their vignettes visually, rather than the
498 fiscal costs of doing so.

499 Future efforts should strive to require even less textual communication while participants
500 are answering questions. Illustrating a social scenario via our visual vignette method should
501 always require far less reading than a textual vignette, but researchers can employ creative
502 methods to further reduce the amount of verbal skill necessary. Generic sliders representing good
503 and bad could be applied to populations who speak different languages and research has shown
504 that good/bad evaluations tend to convey meaning universally across cultures (Osgood 1962).
505 Using simple dimensions of meaning would make it much easier to accurately translate brief
506 survey items and instructions than to translate vignettes without introducing artifacts. Future
507 studies should use our visual vignette method to test how different cultures evaluate the same
508 social situation. As we were interested in validating the basic utility of our approach, it made
509 sense to do so with a single national population. But there is no reason why our method could not
510 be used to obtain equivalent interaction data from multiple nations in a larger effort. Our
511 vignettes rely on simple geometric shapes with minimal humanizing characteristics and
512 therefore, unlike human actors, only contain the features that researchers want them to contain.
513 Future studies should explore the limits of this approach, both by introducing analogs for human
514 features of interest (e.g., secondary sexual characteristics) and introducing distinctions analogous
515 to human status characteristics but without existing status value (e.g., blue and orange “skinned”
516 characters). This latter approach in particular offers the opportunity to disentangle “mere

517 difference” effects from those of established beliefs about the meaning of specific physical
518 characteristics.

519 **Conclusion**

520 Our study demonstrates how a visual vignette approach that uses minimally human-like
521 characters can produce similar results when compared to more traditional textual vignettes.

522 While we failed to find gender effects for the characters in our vignettes, our results suggest that
523 visual vignettes not only are a viable alternative to textual vignettes, but they could produce more
524 potent effects as well. This method has considerable promise to improve research both with
525 cross-national populations, by ensuring that the stimulus presented is truly common, and with
526 low literacy populations. Moreover, it is flexible enough to be integrated into both online data
527 collection, as with our study, or to be deployed in the field as a part of face-to-face interviewing
528 (using laptop or tablet computers). By harnessing the human tendency to view even non-social
529 events in social terms, we gain an improved ability to understand the interactions between
530 humans themselves.

531

532

533

534

535

536

537

References

- 538
- 539 Aguinis, H., & Bradley K.J. (2014). Best practice recommendations for designing and
540 implementing experimental vignette methodology studies. *Organizational research*
541 *methods* 17, 4, 351-371.
- 542 Alves, W. M., & Rossi, P. H. (1978). Who should get what? Fairness judgments of the
543 distribution of earnings. *American journal of Sociology*, 84(3), 541-564.
- 544 Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior:
545 Judgmental accuracy from thin slices of the behavioral stream. In *Advances in*
546 *experimental social psychology* (Vol. 32, pp. 201-271). Academic Press.
- 547 Anast Seguin, C., & Ambrosio, A. L. (2002). Multicultural vignettes for teacher
548 preparation. *Multicultural Perspectives*, 4(4), 10-16.
- 549 Aries, E. (1987). *Gender and communication*. Sage Publications, Inc.
- 550 Atzmüller, C., and Steiner, P.M. (2010). Experimental vignette studies in survey research.
551 *Methodology*, 6(3):128–138.
- 552 Balslev, T., De Grave, W. S., Muijtjens, A. M., & Scherpbier, A. J. J. A. (2005). Comparison of
553 text and video cases in a postgraduate problem-based learning format. *Medical*
554 *education*, 39(11), 1086-1092.
- 555 Barrera, D and Buskens, V. (2002). Imitation and learning under uncertainty: A vignette
556 experiment. *ICS/Department of Sociology, Utrecht University*.
- 557 Baron-Cohen, S. (2000). Theory of mind and autism: A fifteen year review. *Understanding other*
558 *minds: Perspectives from developmental cognitive neuroscience* 2, 3-20.

559 Berger, J., Webster, M., Ridgeway, C., and Rosenholtz, S.J (2018). Status cues, expectations,
560 and behavior. In Morris Zelditch, *Status, Power, and Legitimacy*, 155-174. Routledge.

561 Bongiorno, R, Bain, P.G. and David, B. (2014). If you're going to be a leader, at least act like it!
562 Prejudice towards women who are tentative in leader roles. *British Journal of Social*
563 *Psychology* 53, 2, 217-234.

564 Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of Amazon's
565 Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological*
566 *Science*, 13(2), 149-154.

567 Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants
568 and data gathered via Amazon's MTurk, social media, and face-to-face behavioral
569 testing. *Computers in human behavior*, 29(6), 2156-2160.

570 Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: a functional imaging
571 study of perception and interpretation of complex intentional movement
572 patterns. *Neuroimage*, 12(3), 314-325.

573 De Leng, B. A., Dolmans, D. H., Van de Wiel, M. W., Muijtjens, A. M. M., & Van Der Vleuten,
574 C. P. (2007). How video cases should be used as authentic stimuli in problem-based
575 medical education. *Medical education*, 41(2), 181-188.

576 Difallah, D., Filatova, E., & Ipeirotis, P. (2018, February). Demographics and dynamics of
577 mechanical Turk workers. In *Proceedings of the eleventh ACM international conference*
578 *on web search and data mining* (pp. 135-143).

579 Dunbar, R. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and*
580 *Reviews: Issues, News, and Reviews* 6, 5, 178-190.

581 Farley, S. D. (2008). Attaining status at the expense of likeability: pilfering power through
582 conversational interruption. *Journal of Nonverbal Behavior* 32, 4, 241-260.

583 Foschi, M. (1992). Gender and double standards for competence. In Cecilia Ridgeway, *Gender,*
584 *Interaction, and Inequality*, 181-207. New York: Springer.

585 Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal
586 infants. *Nature*, 450(7169), 557-559.

587 Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. P. (2018, April).
588 A data-driven analysis of workers' earnings on Amazon Mechanical Turk. In *Proceedings*
589 *of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

590 Hare, A.P, and Bales R.F. (1963). Seating position and small group interaction. *Sociometry*, 480-
591 486.

592 Heflick, N. A., Goldenberg, J. L., Cooper, D. P., & Puvia, E. (2011). From women to objects:
593 Appearance focus, target gender, and perceptions of warmth, morality and
594 competence. *Journal of Experimental Social Psychology*, 47(3), 572-581.

595 Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American*
596 *journal of psychology*, 57(2), 243-259.

597 Heise, D. R. (2007). *Expressive order: Confirming sentiments in social actions*. Springer Science
598 & Business Media.

599 Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not
600 WEIRD. *Nature*, 466(7302), 29-29.

601 Heir, W. (2000). Minimizing workplace gender and racial bias. *Psychology* 37, 233-56.

602 Hughes, R. (1998). Considering the vignette technique and its application to a study of drug
603 injecting and HIV risk and safer behaviour. *Sociology of Health & Illness* 20, 3, 381-400.

604 Janicik, G.A., and Larrick R.P. (2005). Social network schemas and the learning of incomplete
605 networks. *Journal of personality and social psychology* 88, 2, 348-364.

606 Kelemen, D. (1999). Function, goals and intention: Children's teleological reasoning about
607 objects. *Trends in Cognitive Sciences* 3, 12, 461-468.

608 Klin, A. (2000). Attributing social meaning to ambiguous visual stimuli in higher-functioning
609 autism and Asperger syndrome: The social attribution task. *The Journal of Child*
610 *Psychology and Psychiatry and Allied Disciplines* 41, 7, 831-846.

611 LaFrance, M. (1992). Gender and interruptions: Individual infraction or violation of the social
612 order?. *Psychology of Women Quarterly* 16, 4, 497-512.

613 Lott, D.F., and Sommer R. (1967). "Seating arrangements and status." *Journal of personality and*
614 *social psychology* 7, 1, 90-95.

615 Lovett, M., Bajaba, S., Lovett, M., & Simmering, M. J. (2018). Data quality from crowdsourced
616 surveys: A mixed method inquiry into perceptions of amazon's mechanical turk
617 masters. *Applied Psychology*, 67(2), 339-366.

618 Marlowe, C. M., Schneider, S. L., & Nelson, C. E. (1996). Gender and attractiveness biases in
619 hiring decisions: Are more experienced managers less biased?. *Journal of applied*
620 *psychology, 81*(1), 11.

621 Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A
622 quantitative cartography of the Uncanny Valley. *Cognition, 146*, 22-32.

623 Nemeth, C., & Wachtler, J. (1974). Creating the perceptions of consistency and confidence: A
624 necessary condition for minority influence. *Sociometry, 529-540*.

625 O'Connor, K. M., & Gladstone, E. (2015). How social exclusion distorts social network
626 perceptions. *Social Networks, 40*, 123-128.

627 Osgood, C.E. (1962). Studies on the generality of affective meaning systems. *American*
628 *Psychologist 17*, 1, 10-28.

629 Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality
630 on Amazon Mechanical Turk. *Behavior research methods, 46*(4), 1023-1031.

631 Simon, M., and Tierney. R. (2011). Use of vignettes in educational research on sensitive teaching
632 functions such as assessment. In *Wystapienie na: 21th International Congress for School*
633 *Effectiveness and Improvement*.

634 Simpson, B., Markovsky, B., & Steketee, M. (2011). Power and the perception of social
635 networks. *Social Networks, 33*(2), 166-171.

636 Slead, M., Durrheim, K., Kriel, A., Solomon, V., & Baxter, V. (2002). The effectiveness of the
637 vignette methodology: A comparison of written and video vignettes in eliciting responses
638 about date rape. *South African Journal of Psychology, 32*(3), 21-28.

639 Steckenfinger, S. A., & Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny
640 valley. *Proceedings of the National Academy of Sciences*, 106(43), 18362-18366.

641 Thye, S. R., & Harrell, A. (2017). The status value theory of power and mechanisms of micro
642 stratification: Theory and new experimental evidence. *Social science research*, 63, 54-66.

643 Tippett, E. C. (2018). Harassment trainings: A content analysis. *Berkeley J. Employment. &*
644 *Labor Law*. 39, 481-526.

645 Wason, K. D., Polonsky, M. J., & Hyman, M. R. (2002). Designing vignette studies in
646 marketing. *Australasian Marketing Journal (AMJ)*, 10(3), 41-58.

647 Woehr, David J., and Charles E. Lance. 1991. "Paper people versus direct observation: An
648 empirical examination of laboratory methodologies." *Journal of Organizational*
649 *Behavior* 12, 5, 387-397.

650 Whiten, A., and Byrne R.W., eds. (1997). Machiavellian intelligence II: Extensions and
651 evaluations. Vol. 2. *Cambridge University Press*.

652 Yadav, A., Phillips, M. M., Lundeberg, M. A., Koehler, M. J., Hilden, K., & Dirkin, K. H.
653 (2011). If a picture is worth a thousand words is video worth a million? Differences in
654 affective and cognitive processing of video and text cases. *Journal of Computing in*
655 *Higher Education*, 23(1), 15-37.

656

657

658

659

660

Appendix A

661 ***Textual vignette version 1:***

662 Vision advertising is a successful advertising agency. To promote a new product for one of their
663 clients, they have put together a 4-person team from outside the agency. The team members are:

664

665 Melissa, 42 years old, female;

666 John, 43 years old, male;

667 Cindy, 28 years old, female;

668 David, 26 years old, male.

669

670 They have not decided the responsibilities or who is in charge. Right before their first meeting,
671 Melissa chose to sit at the head position of the table as they entered the conference room. John
672 and Cindy started to talk with each other about graphic designing ideas. During the
673 conversation, John barely looked at Cindy. David wanted to discuss the selling point of the new
674 product and interrupted the conversation, cutting off Cindy. Cindy then turned away to talk
675 with Melissa.

676

677 ***Textual vignette version 2:***

678 Vision advertising is a successful advertising agency. To promote a new product for one of
679 their clients, they have put together a 4-person team from outside the agency. The team
680 members are:

681

682 Melissa, 42 years old, female;

683 John, 43 years old, male;

684 Cindy, 28 years old, female;

685 David, 26 years old, male.

686

687 They have not decided the responsibilities or who is in charge. Right before their first
688 meeting, John chose to sit at the head position of the table as they entered the conference
689 room. Melissa and David started to talk with each other about graphic designing ideas.
690 During the conversation, Melissa barely looked at David. Cindy wanted to discuss the selling
691 point of the new product and interrupted the conversation, cutting off David. David then
692 turned away to talk with John.

693

694