

category. Unwarranted inferences are the result of speculation about the image; here, the annotator goes beyond what can be glanced from the image and makes use of their knowledge and expectations about the world to provide an overly specific description. Such descriptions are directly identifiable as such, and in fact we have already seen four of them (descriptions 2–5) discussed earlier.

2.1. Linguistic bias

Generally speaking, people tend to use more concrete or specific language when they have to describe a person that does not meet their expectations. Beukeboom (2014) lists several linguistic ‘tools’ that people use to mark individuals who deviate from the norm. I will mention two of them.²

Adjectives One well-studied example (Stahlberg et al., 2007; Romaine, 2001) is sexist language, where the sex of a person tends to be mentioned more frequently if their role or occupation is inconsistent with ‘traditional’ gender roles (e.g. *female surgeon*, *male nurse*). Beukeboom also notes that adjectives are used to create “more narrow labels [or subtypes] for individuals who do not fit with general social category expectations” (p. 3). E.g. *tough woman* makes an exception to the ‘rule’ that women aren’t considered to be tough.

Negation can be used when prior beliefs about a particular social category are violated, e.g. *The garbage man was not stupid*. See also (Beukeboom et al., 2010).

These examples are similar in that the speaker has to put in additional effort to mark the subject for being unusual. But they differ in what we can conclude about the speaker, especially in the context of the Flickr30K data. Negations are much more overtly displaying the annotator’s prior beliefs. When one annotator writes that *A little boy is eating pie without utensils* (image 2659046789), this immediately reveals the annotator’s normative beliefs about the world: pie should be eaten *with* utensils. But when another annotator talks about *a girls basketball game* (image 8245366095), this cannot be taken as an indication that the annotator is biased about the gender of basketball players; they might just be helpful by providing a detailed description. In section 3 I will discuss how to establish whether or not there is any bias in the data regarding the use of adjectives.

2.2. Unwarranted inferences

Unwarranted inferences are statements about the subject(s) of an image that go beyond what the visual data alone can tell us. They are based on additional assumptions about the world. After inspecting a subset of the Flickr30K data, I have grouped these inferences into six categories (image examples between parentheses):

Activity We’ve seen an example of this in the introduction, where the ‘manager’ was said to be *talking about job performance* and *scolding [a worker] in a stern lecture* (8063007).

Ethnicity Many dark-skinned individuals are called *African-American* regardless of whether the picture has been taken in the USA or not (4280272). And people who



Figure 2: Image 4183120 from the Flickr30K dataset.

look Asian are called Chinese (1434151732) or Japanese (4834664666).

Event In image 4183120 (Figure 2), people sitting at a gym are said to be watching a game, even though there could be any sort of event going on. But since the location is so strongly associated with sports, crowdworkers readily make the assumption.

Goal Quite a few annotations focus on explaining the *why* of the situation. For example, in image 3963038375 a man is fastening his climbing harness *in order to have some fun*. And in an extreme case, one annotator writes about a picture of a dancing woman that *the school is having a special event in order to show the american culture on how other cultures are dealt with in parties* (3636329461). This is reminiscent of the Stereotypic Explanatory Bias (Sekaquaptewa et al., 2003, SEB), which refers to “the tendency to provide relatively more explanations in descriptions of stereotype inconsistent, compared to consistent behavior” (Beukeboom et al., 2010, p. 5). So in theory, odd or surprising situations should receive more explanations, since a description alone may not make enough sense in those cases, but it is beyond the scope of this paper to test whether or not the Flickr30K data suffers from the SEB.

Relation Older people with children around them are commonly seen as parents (5287405), small children as siblings (205842), men and women as lovers (4429660), groups of young people as friends (36979).

Status/occupation Annotators will often guess the status or occupation of people in an image. Sometimes these guesses are relatively general (e.g. college-aged people being called *students* in image 36979), but other times these are very specific (e.g. a man in a workshop being called a *graphics designer*, 5867606).

3. Detecting stereotype-driven descriptions

In order to get an idea of the kinds of stereotype-driven descriptions that are in the Flickr30K dataset, I made a browser-based annotation tool that shows both the images and their associated descriptions.³ You can simply leaf through the images by clicking ‘Next’ or ‘Random’ until you find an interesting pattern.

²Examples given are also due to (Beukeboom, 2014).

³Code and data is available on GitHub: <https://github.com/evanmiltenburg/Flickr30K-Image-Viewer>