

# How much truth is in stereotypes?

Szymon Czarnik

Institute of Sociology, Jagiellonian University

## Abstract

Stereotype accuracy is a contentious topic. Part of the problem is that typically stereotypes are generic statements whose truth status is unclear due to the fact that they are ill-defined quantitatively. The article focuses on the epistemic aspect of stereotypical beliefs. In the ongoing debate, I side with those who argue against stereotypes being wrong or inaccurate by virtue of definition alone. I propose that, when possible, stereotype accuracy should be assessed in probabilistic terms by inspecting how likely a generic statement is to be true when applied to individual(s) representative of the relevant group(s). This approach applies equally well to investigating the actual and the perceived accuracy of stereotypes.

## Keywords

stereotypes, stereotype accuracy, epistemology, probabilistic approach.

**I**n this paper, I propose that stereotypes are best understood as generic statements whose truth status is unclear. They are expressed by asserting that “members of group A have trait T,” or “Members of group A are higher/lower on trait T than members of group B.” Such utterances lack proper quantification and invite alternative interpretations. One such interpretation is that stereotypes

speak of all group members without exception which leads to the idea that stereotypes should be considered false by definition, which in turn leads to the opposition to stereotype accuracy research as, at best, a misguided endeavor. This opposition, however, is unfounded on epistemic grounds once we explicitly recognize generic nature of stereotypes. Thus after outlining the long-lasting debate, I proceed to discuss the problems inherent in measuring the accuracy of stereotypes, and then advocate a probabilistic approach, i.e. assessment of probabilities that individuals conform to the generic rule posited by a stereotype. As a practical illustration, I show a number of such probabilistic assessments based on data from representative samples. Finally, I propose that the very same probabilistic approach should be used to directly investigate how accurate people perceive their own stereotypes to be. This in turn would provide an easy and straightforward method of comparing perceived and actual accuracy of stereotypes, acknowledging a possibility that stereotypical perceptions may both over- and underestimate actual differences.

\* \* \*

A scholarly dispute over stereotypes is to no small extent a semantic issue. The problem is that a typical stereotype assumes a form of what *The Stanford Encyclopedia of Philosophy* calls a generic, i.e. a statement which expresses a generalization “but unlike quantified statements, [...] do[es] not carry information about how many members of the kind or category have the property” (Leslie and Lerner, 2016). “Some men are taller than some women” is certainly a true statement. Equally certainly, “All men are taller than all women” is a false statement. But the case is that when you hear anybody compare the height of men and women, it will be in the form of a generic

“Men are taller than women.” A quantitative indeterminacy of such an utterance will either be construed to be vaguely true and will be silently approved, or will invite a dispute and be contested.

In daily life conversations, calling somebody’s remark a stereotype is surely a form of criticism. In common parlance, as well as in vast sociological and psychological literature, the term ‘stereotype’ implies untruth, or inaccuracy of the assertion. A general statement like “Scots are cheap” may be picked apart as untrue in the sense that on average Scots are no more stingy than members of any other ethnic group. Or, even if the statement had some kernel of truth in it, one could be brought to task for inaccurately applying it to pass a judgment on a particular individual who may happen not to conform to this general rule.<sup>1</sup> Apart from being misguided epistemologically, the utterance may also be criticized on moral grounds as an attempt at a wholesale vilification of a certain group of people. Thus the presumed wrongness of stereotypes may consist in epistemic and/or moral failure (Beeghly, 2014). It is solely the former aspect of stereotypes that we shall be concerned with here.<sup>2</sup>

## **1. Resistance to research on stereotype accuracy**

Before we proceed further it should be acknowledged that, for a good reason, the ethical aspect looms large in the scholarly literature on

---

<sup>1</sup> Far from being limited to stereotypical beliefs, the problem of group-to-individual generalizability may pose a serious threat to scientific research involving human subjects (Fisher, Medaglia and Jeronimus, 2018).

<sup>2</sup> Some authors reduce stereotypes to the epistemic aspect alone. Thus Banaji claims that “[f]rom the earliest use of the term in psychology, stereotypes have been regarded as the cognitive (thought) as opposed to the affective (feeling) component of mental representations of social groups. As such, the construct is tied to but differentiated from the concept of attitude, preference, or liking as well as the concept of discrimination” (Banaji, 2001, p.15101).

the topic. In psychology, stereotypes are studied as cognitive components of prejudice, which may be brought about by and contribute to inter-group hostility (Bar-Tal, 1989); in sociology, they are recognized as a device that may adversely affect targeted groups by means of self-fulfilling prophecy (Merton, 1948); in criminology, they are discussed in the context of racial/ethnic profiling in police operations (Schauer, 2003); in historical studies, they are investigated as part and parcel of totalitarian propaganda (Werth, 2011; USHMM (United States Holocaust Memorial Museum), n.d.); in economics, they are incorporated in models of statistical discrimination in the labor market (Arrow, 1971; Phelps, 1972). It goes without saying that after Adorno had linked stereotypic thinking with fascist belief systems the concept became even more anathema (Jones and Colman, 2001).

It is then easy to see why the very idea of considering stereotype accuracy has been frowned upon. Being a lively research field until around mid-1950s, in later years accuracy research nearly ceased to exist (Jussim, 2012). Recognized harmful implications of stereotypes made most researchers extremely reluctant to lend them any credibility for fear of promoting any sort of prejudice. In most of the writings on stereotypes, this reluctance tended to be implicitly transformed into a non-sequitur assumption that “if stereotypes are associated with social wrongs, they must be factually wrong” (Jussim, Cain et al., 2009, p.199). This conflation was brought up and chastised as early as 1965 by Roger Brown who opined that it was a disingenuous device by means of which the social psychologist “perverted his science to achieve a moral purpose” (Stroebe and Insko, 1989, p.5).<sup>3</sup>

---

<sup>3</sup> We encounter a similar sentiment even earlier in a seminal book (Allport, 1954, p.87): “we find social scientists who overhastily reject the very possibility of racial, national, or group differences of any appreciable or fundamental order. Some of them do so on the basis of charitable motives, but the evidence they offer is usually fragmentary”. He even goes as far as to urge researchers to investigate “all the facts we

Still thirty years later, going against the tide, the editors of *Stereotype Accuracy. Toward Appreciating Group Differences* forewarned in the preface:

It is not easy to do research on stereotype accuracy, for both scientific and political reasons [...]. The intellectual content of this book commits multiple heresies [...] the idea that stereotypes may sometimes have some degree of accuracy is apparently anathema to many social scientists and laypeople. Those who document accuracy run the risk of being seen as racists, sexists, or worse (Lee, Jussim and McCauley, 1995, p.xiii).<sup>4</sup>

Needless to say, authors go to great lengths to explicate why the slur is utterly unfounded.

## 2. Ambiguity about truth status of stereotypes

Hostility toward stereotype accuracy reported by Jussim et al. notwithstanding, one should recognize that from the very outset there was a lot of ambiguity about the truth status of stereotypes. Two of the most seminal pieces on stereotypes, viz. Part III of *Public Opinion* by Walter Lippmann (credited with introducing the concept in

---

can get to evaluate the alleged claim that a hated group merits hostility—that its evil reputation is well deserved” (Allport, 1954, p.104)—an utterance that in the present day and age would likely bring upon him a charge of “blaming the victim”.

<sup>4</sup> Indeed, stereotypes acquired such a contemptible connotation that it became a useful strategy for overbearing intellectuals to attach the term to any commonsensical statement they found unpalatable. As Thomas Sowell observed, “the very conception of testing beliefs against reality is attacked by such things as deconstruction, cultural relativism, and the practice of describing uncongenial conclusions as ‘perceptions’ or ‘stereotypes’ and attributing ‘false consciousness’ to those who hold them” (Sowell, 1995, p.244).

1922), and *The Nature of Prejudice* by Gordon W. Allport (1954) are case in point. They both consider stereotype formation as a natural and inevitable process taking place in human mind when people try to make sense of a limited amount of information. Likewise, they both admit that beliefs about groups may be correct to different degrees. Thus, for example, Lippmann (1998, p.90) observes:

Were there no practical uniformities in the environment, there would be no economy and only error in the human habit of accepting foresight for sight.<sup>5</sup> But there are uniformities sufficiently accurate, and the need of economizing attention is so inevitable, that the abandonment of all stereotypes for a wholly innocent approach to experience would impoverish human life.

But having admitted this much, he offers a very telling example of what he considers to be a perfect stereotype. He invokes Aristotle's infamous defense of slavery in Book I of *Politics*. Therein the philosopher instructs his readers that nature created the bodies of slaves and free men different from each other: the former fit for servile labor, the latter for civil life. Utter nonsense as it is, notes Lippmann, training Greeks to see slaves as slaves by nature was presumably the only "righteous" justification of the institution one could offer. He then sums up Aristotle's persuasion strategy:

Each slave holder was to look upon his chattels as natural slaves. When his eye had been trained to see them that way, he was to note as confirmation of their servile character the fact that they performed servile work, that they were competent to do servile work, and that they had the muscles to do servile work.

---

<sup>5</sup> Lippmann construed stereotypes as preconceptions that guide and organize our perception.

This is the perfect stereotype. Its hallmark is that it precedes the use of reason; is a form of perception, imposes a certain character on the data of our senses before the data reach the intelligence [...]. There is nothing so obdurate to education or to criticism as the stereotype. (p.98)

In a similar fashion Allport (1954, pp.189–190), while conceding that some stereotypes “need not be altogether false”, points to the fact that there are as well some that “are totally unsupported by facts” and it’s possible for them “to grow in defiance of all evidence.”

### 3. Are stereotypes untrue by definition?

There is not one agreed upon definition of stereotypes. Probably the shortest and most general definition of stereotypes as “category-based knowledge” was provided *en passant* by Macrae and Quadflieg in a chapter on perceiving people (Macrae and Quadflieg, 2010, p.428). In the social context, such categories would predominantly be groups of people sharing a certain salient characteristic. Naturally enough, most definitions add certain qualifications, of which the one pertinent to our topic is whether stereotypes are necessarily untrue, or inaccurate. As Stroebe and Insko (1989) point out, stereotypes have often been defined as “incorrect generalizations”; they were variously meant to be “relatively unresponsive to external reality” (Klineberg, 1951), “exaggerated” (Allport, 1954), “biased” (English and English, 1958), “generally invalid” (Miller and Turnbull, 1986)<sup>6</sup>, “oversim-

---

<sup>6</sup> It is ironic and telling that defining stereotypes as generally invalid, Miller and Turnbull quote Ashmore and Del Boca (1981) who, in the very paper quoted, explicitly advocated accuracy-neutral definition: “Stereotypes have been defined as ‘bad’ for one or a combination of the following reasons: A stereotype is a set of beliefs that

plified” (Koleser, 2008), “over-generalized” (Wikipedia, 2020), etc.<sup>7</sup> However, Jussim et al. (2009) supplied a cogent argument against such defining away of stereotype’s accuracy. The argument is two-fold as, logically, there are only two possible interpretations of the assertion that all stereotypes are inaccurate. First reading is that all beliefs about groups are stereotypes and they are all wrong. As a matter of fact, hardly anybody subscribes to such a claim. It is quite obviously self-defeating because if it was true, any meaningful discussion about groups of people would *ipso facto* be impossible. Second interpretation is that not all beliefs about groups are wrong and we call stereotypes only those that are.<sup>8</sup> The problem with this approach is that it is inconsistent with the actual practice of research on stereotypes. It necessitates that the first step in stereotype research should be to demonstrate that the belief under investigation is indeed factually wrong. Otherwise, it would plainly not be a stereotype study. However, in practice that first step is typically skipped, the burden of proof is left untouched and it is simply assumed that the belief is

---

is incorrectly learned, overgeneralized, factually incorrect, or rigid [...]. While stereotypes may well have any or all of these characteristics, the proposed sources of ‘badness’ should not be incorporated into the definition of the term ‘stereotype’” (Miller and Turnbull, 1986, p.16).

<sup>7</sup> Other characteristics, frequently added, have rendered stereotypes negative, resistant to change, and disindividualized (shared within a particular social environment).

<sup>8</sup> This in fact was the position taken by Allport who made an explicit distinction between stereotypes (distorted beliefs) and group traits (real differences between groups): “If [an image of a category] is a generalized judgment based on a certain probability that an object of the class will possess a given attribute, we would not call the judgment a stereotype” (Allport, 1954, p.189). And then: “We can distinguish between a valid generalization and a stereotype only if we have solid data concerning the existence of (the probability of) true group differences” (Allport, 1954, p.192).



false.<sup>9</sup> Jussim et al. (2016, p.32) dub this phenomenon “The Black Hole at the Bottom of Many Declarations of Stereotype Inaccuracy”. By means of exemplification, they quote a passage from a world-renowned social psychology handbook stating that “[t]o stereotype is to allow those pictures [i.e. our mental images of external reality] to dominate our thinking, leading us to assign identical characteristics to any person in a group, regardless of the actual variation among members of that group” (Aronson, 2011, p.309). As they are quick to notice, Aronson “does not cite anything to support such an extreme claim because he cannot. After nearly 100 years of empirical research on social stereotypes, there is not a single study that has reported a single person who believes all members of any group have identical characteristics” (Jussim, Crawford, Anglin et al., 2016, p.33).

Still another more serious problem is that stereotype itself is a categorical concept, and since beliefs about groups may be located anywhere on the continuum between blatant falsehood and clear truth, to categorize a statement as a stereotype one needs to choose a cut-off point—a purely arbitrary decision. To be sure, any academic might argue that the cut-off point should simply be 100% true; that a given stereotype, e.g. that men are taller than women, may be considered true only if each and every man is taller than any woman.<sup>10</sup> But then one is left wondering how many of the propositions based on social-scientific studies would hold their ground for more than five seconds when faced with such an absurdly high standard of truth.<sup>11</sup>

---

<sup>9</sup> The reason for this is that “social psychologists once so firmly believed in stereotype inaccuracy, that declaring stereotypes to be inaccurate did not even require a reference!” (Jussim, Crawford, Anglin et al., 2016, p.33)

<sup>10</sup> As the above quote from Aronson shows, this line of reasoning is routinely employed to demonstrate that stereotypes are false.

<sup>11</sup> As a matter of fact, Jussim et al. (2016) make a point of demonstrating that correlations between stereotypic beliefs and actual data on group characteristics are substantially higher than those observed in most studies in social psychology.

As it is, it stands to reason that this Gordian knot should be cut—as Jussim et al. propose—by agreeing on an accuracy-neutral definition of stereotype, making it an umbrella term for all beliefs about groups, and leaving it open to empirical investigation how much trust we can put in any of them.<sup>12</sup> Obviously, as they explicitly make clear, “rejection of defining stereotypes as inaccurate is not equivalent to defining them as accurate.” (Jussim, Crawford, Anglin et al., 2016, p.202). A good example of an accuracy-neutral definition is the one provided by Ashmore and Del Boca (1981, p.16), wherein the core meaning of a stereotype (i.e. shared by all scholars dealing with the subject) makes it simply a “set of beliefs about the personal attributes of a group of people”. This definition is then followed by a series of arguments why stereotypes should not be treated as intrinsically “bad”.

#### **4. Problems in measuring the accuracy of stereotypes<sup>13</sup>**

Assessing stereotype (in)accuracy involves three obvious steps (Jussim, Crawford and Rubinstein, 2015):

---

<sup>12</sup> For the very same reason, it would be wise not to define stereotypes as a type of commonly held beliefs. Again, the commonness of a particular stereotype could be determined empirically. At one extreme, one could even speak of a single person having a peculiar stereotype of a given group. Thus Jussim et al. (2016) distinguish between individual and consensual stereotypes. Apart from that, the problem of measuring belief’s accuracy is completely independent of how many people really believe it.

<sup>13</sup> It should be stressed that what I am interested in here is the extent to which stereotypical judgments as such are correct. It is a related but separate issue whether the process by which people arrive at their judgments is correct. For example, someone may arrive at a correct estimate of honesty in a given group of people, even though he overestimates honesty of people in general while underestimating honesty of that particular group relative to general public. See (Jussim, Crawford, Anglin et al., 2016) for a discussion of processual models of judgment by Cronbach (1955) and Kenny (1994).

1. Eliciting person's belief about certain group(s) with regard to a particular characteristic.
2. Identifying criteria, i.e. relevant reference data on the target group.
3. Comparing beliefs to criteria.

Simple as it sounds, the whole enterprise is fraught with difficulties and this is why some authors would give up on the idea of accuracy measurement altogether.<sup>14</sup> First of all, the meaning of the stereotype will often be ambiguous. Consider the claim that men make better drivers than women. What does it actually mean? Is it about them being faster in moving from A to B? More observant of traffic rules? Having higher parking skills? Having caused fewer accidents per kilometer travelled? To be sure, all these things should count in a good driver competition yet different questions are likely to yield different answers. To overcome this problem, one needs either to narrow down the operational definition of "goodness" in driving, or alternatively, construct an index encompassing all the relevant qualities.

Secondly, one may be hard-pressed to convincingly operationalize the trait. While some traits are easily measurable (e.g. body height and weight), other are complex constructs, though more or less straightforwardly linkable to the existing data (e.g. criminality), and still other are elusive enough to trigger off a dispute, whichever way they would be operationalized (e.g. egoism).

Thirdly, one has to deal with the evaluative component of stereotypes. Some traits are inherently positive (trustworthy, responsible),

---

<sup>14</sup> On the other hand, as (Jussim, Crawford, Anglin et al., 2016) rightly observe, one cannot claim that it is impossible to assess how (in)accurate stereotypes are, and at the same time demand that they are inaccurate.

some negative (cruel, duplicitous), but some may be both, depending on how you choose to frame them (fearless vs. reckless, self-confident vs. arrogant). As Robert Merton suggested, the very same personality qualities, though differently labeled, may be quoted to extol Abraham Lincoln, and deplore Jews. He was thrifty while they are tight-fisted, he was ambitious while they are pushing, he was devoted to human rights while they are outright radical (Allport, 1954, p.189). Furthermore, if the inherent evaluation makes the stereotype overtly prescriptive in form (“girls should wear skirts”), one is entirely precluded from assessing its (in)accuracy (Jussim, Crawford, Anglin et al., 2016).

And then, finally, the ultimate obstacle to the measurement may lie in the unavailability of any relevant data whatsoever.

## **5. Types of stereotypes to be assessed for accuracy**

Restricting our analysis to the situations where measurement is possible, we should not expect there to be a universal method of quantifying stereotypes’ accuracy. The optimal strategy will likely depend on the type of stereotype under study, as determined at least by the level of measurement allowable for a particular trait (binary, multinomial, ordinal, interval, or ratio) and the nature of comparison involved. The latter requires some elaboration. Provisionally, I would propose that stereotypes may be grouped into four distinct classes as summarized in Table 1.

Type	Specification	Example	Epistemic consequence(s)
monotomous <sup>15</sup> (one-to-none)	focuses on a single group, no comparison to other groups is implied	Catholics don't eat meat on Fridays.	We assume to know the trait prevalence/intensity in the group.
dichotomous (one-to-one)	there are two relevant groups, one is compared to the other	Women are more law-abiding than men.	We assume to know the relative prevalence/intensity of the trait in both groups.
polytomous (one-to-many)	there are many relevant groups, one is singled out for inspection	Poles make brave soldiers.	We assume to know that the prevalence/intensity of the trait in the focal group is above average. We don't necessarily assume to know anything about any other group in particular.
incremental	groups are implicitly defined and ordered by intensity of certain characteristic	The older, the more clumsy.	We assume to know that prevalence/intensity of the trait is monotonously linked to the group-defining metric.

Table 1: Types of stereotypes

Let us shortly review the types.

1. **Monotomous**, or **one-to-none** stereotypes, are relatively easy to investigate (given the trait can be satisfyingly operationalized, of course). In the “Catholics don’t eat meat on Fridays” example, a straightforward way to measure the amount of truth in the statement would be to find out what is the actual percentage of Catholics who observe this practice. No comparison is necessary here because even if it turned out that other religious denominations (e.g. Orthodox Christians) go vegetarian on Fridays as well, it would not invalidate the stereotype.
2. **Dichotomous**, or **one-to-one** stereotypes, may be formulated in two ways, one being the reverse of the other. If women are more law-abiding than men then men must be less law-abiding than women. The reality check involves the comparison of the two distributions.
3. **Polytomous**, or **one-to-many** stereotypes, are more problematic to assess. If Poles make brave soldiers,<sup>16</sup> they are somehow above the average on the bravery scale but it remains unclear what the timeframe is and who the reference group is

---

<sup>15</sup> This is a kind of stereotype that Allport (1954, pp.97–98) discusses under the headline “J-curve of conformity behavior” (“J” refers to a line on a chart showing very few people low on the trait, slightly more in the middle, and most very high): “The characteristic thing about the J-curve is that only the members of a given group can be fitted to it. It is simply not applicable to nonmembers [...]. Catholics will fit the J-curve of attendance at Mass, but not non-Catholics [...]. The logic of the J-curve, then, may be stated as follows: Whenever there is a strongly prescribed action for members of an ingroup they will, by virtue of their membership, tend to conform [...]. But] J-curves of conformity can *decay*. When fewer and fewer members perform the prescribed actions the distinctive character of the group gradually disappears.”

<sup>16</sup> Admittedly, this is a kind of stereotype that would be exceedingly difficult to operationalize.

comprised of.<sup>17</sup> Even if the composition of the reference group is known, the stereotype does not tell us anything about the bravery of any particular nationality included in it. It may well be the case that certain ethnic group, e.g. the Finns, are even braver on the battlefield than Poles. To make matters worse, the outcome of the comparison may be crucially affected not only by which groups are included, but also by how they are categorized,<sup>18</sup> and whether the reference data for the amalgamate of “other groups” should be weighted by their respective sizes, or not.

4. **Incremental** stereotypes would be tested for accuracy by means of correlation between the trait and the (at least ordinal) group-defining characteristic. It is a moot point whether the actual shape of the relationship should affect the assessment of accuracy. For example, the relationship between age and clumsiness may not be linear at all but rather be depicted by the kind of hockey stick graph—being quite flat until the age of, say, 70 and only then rising perceptibly.

For the sake of simplicity, in the following I will deal with dichotomous stereotypes addressing ordinal or interval/ratio traits. The one-to-one stereotypes are of particular interest, even if I am bound to disagree with Pickering’s notion that “stereotyping is a sign of power [...] that] always operates via strict demarcations between ‘us’ and ‘them’” (Pickering, 2011, p.616).

---

<sup>17</sup> For polytomous stereotypes, McCauley et al. (1980) proposed the reference group to be “humans in general”—a device called into question by Ashmore and Del Boca (1981).

<sup>18</sup> For example, Finns could enter the comparison mingled with Swedes and Norwegians as a single “Nordic nations” category.

## 6. How to measure accuracy of a stereotype?

Two measurement procedures have been typically used to investigate stereotype's (in)accuracy: one based on discrepancies between the target group's perceived and real positions on the traits in question, the other on correlations between the two (Jussim, Crawford, Anglin et al., 2016). If you believe that in your country half of mothers with children under 7 work full-time when in reality it is 60%, your discrepancy score is 10 percentage points—you do not exactly hit the target but your idea is not that bad, either. If your estimates are as close to reality as this on most other traits considered with regard to mothers, your correlation score should be high. You could then be said to hold a reasonably accurate stereotype of a mother with young kids.

The great advantage of both discrepancy and correlation methods is that they can be used to measure the accuracy of all four types of stereotypes. However, on the downside, it relies solely on the averages—either perceived, or real—of the investigated traits. Thus in registering subjects' perceptions, as well as the reference data, all distributions are reduced to their central tendencies. This is unfortunate—particularly when dealing with dichotomous stereotypes—because on the whole the perceived difference between any two groups regarding certain trait will depend not only on the difference between the group means but on the amount of the within-group variance as well.<sup>19</sup> We easily observe that an average tomato

---

<sup>19</sup> Correspondingly, Lee and Fiske (2008, p.138), following Judd and Park (1993), distinguish *stereotypic inaccuracy*, i.e. the exaggeration or downplaying of the target group's characteristic from *dispersion inaccuracy*, i.e. perceiving the group as more homogenous, or less homogenous than it really is. Both the exaggeration of the difference and the downplaying of the within-group variances will lead to the stronger perceived separation of the two groups.



is heavier than an average strawberry, even though the difference may amount only to 100 g, but the same average difference will be imperceptible when comparing pumpkins and watermelons. The problem with measuring stereotypic perceptions by taking into account both central tendency and dispersion is that even though people have intuitive understanding of the averages, the statistical concept of variance will be foreign to most of them.<sup>20</sup> Fortunately, we can take advantage of the fact that between- and within-group variances combine to produce a probabilistic differential between the two groups.<sup>21</sup>

A simple and straightforward measure of stereotype accuracy is the probability that a generic stereotypical statement is true when considering individuals. For monotomous stereotypes, it would simply be the probability that a randomly selected member of a group conforms to the stereotype. For dichotomous stereotypes of the type “members of group A are higher on trait T than members of group B,” it would be the probability that a randomly selected individual from Group A will indeed be higher on T than a randomly selected member of Group B.<sup>22</sup>

---

<sup>20</sup> This problem may be solved by dividing the scale into a number of categories and then asking subjects about the percentages of the target group members with very low, low, medium, high and very high scores, respectively. This allows the researcher to estimate the subject’s perceived variance of the trait in the target group (e.g. see Swim, 1994). However, the procedure is tedious and its reliability may be a matter of dispute.

<sup>21</sup> Perhaps it should be noted here that certain stereotypes might be rooted not so much in the perception of the general overlap between the groups as in the attention being focused at one particular tail of the distribution. For example, this seems to be the case with “boys are better at math” stereotype with the large sex differential located at the positive extreme of math ability (Benbow and Lubinski, 1993). In such situations, the stereotypic belief will likely fail the proposed probabilistic accuracy test (and other typical accuracy tests as well).

<sup>22</sup> In a similar vein, McCauley et al. (1980) proposed that for dichotomous stereotypes concerning binary traits, stereotype’s strength be measured by “diagnostic ratio,” i.e. the trait probability in the target group divided by the trait probability in the reference group.

This approach capitalizes on the construal of stereotypes as probabilistic statements rather than unrealistic all-or-nothing kind of assertions (Schneider, 2005). It also constitutes a direct link between the two levels of analysis, collective and individual, that often get mixed up in the stereotype debate (Jussim, Cain et al., 2009). The probabilistic assessment of stereotype accuracy addresses both questions simultaneously: what good are stereotypes when we judge groups as wholes, and what good they are when we judge individual members of those groups.

## 7. Probability that a stereotype is true when applied to individuals

For continuous normally distributed traits, application of the probabilistic approach is relatively straightforward. The probability that a randomly chosen member of the target group T will be higher on trait X than a randomly selected member of the reference group R is given by the formula:

$$p(X_T > X_R) = \frac{1}{2} \left( 1 + \operatorname{erf} \left[ \frac{\mu_T - \mu_R}{\sqrt{2(\sigma_T^2 + \sigma_R^2)}} \right] \right),$$

where  $X_T$  and  $X_R$  are randomly selected values of trait X,  $\mu_T$  and  $\mu_R$  are the trait's means, and  $\sigma_T$  and  $\sigma_R$  are the trait's standard deviations for the target and reference group, respectively, while  $\operatorname{erf}()$  is the Gauss error function.

In Figure 1, we assume that in the reference group the trait is normally distributed with  $\mu_R = 0$  and  $\sigma_R = 1$ , and show how  $p(X_T > X_R)$  depends on the parameters of the target group,  $\mu_T$  and

$\sigma_T$  ( $\mu_T > 0$ ). Various possible values of the target’s standard deviation are plotted on the horizontal axis, and various possible values of the target’s mean are each represented by a separate line<sup>23</sup>. To be sure, for any given standard deviation, the more the target’s mean deviates from the reference mean, the higher the probability that  $X_T > X_R$ . And, as evidenced by the downward slope of the lines, for any given difference of the means, the larger the target group’s standard deviation, the lower the probability that  $X_T > X_R$ .

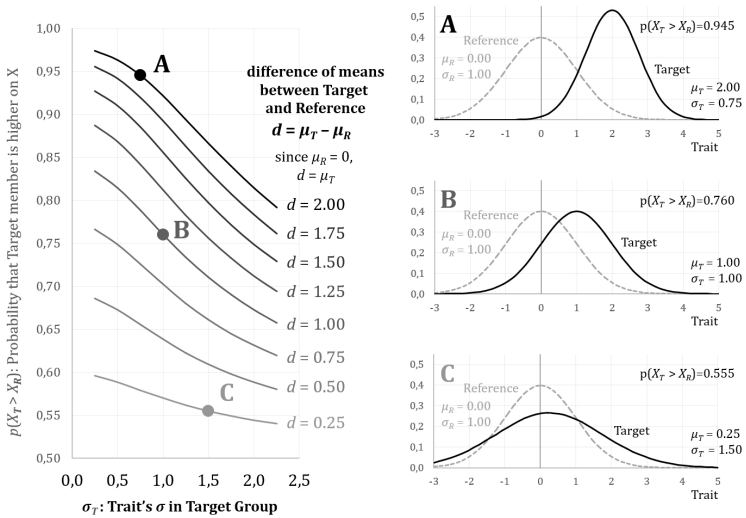


Figure 1: Probability of a stereotype being true depending on the overlap of the trait’s distributions in the target and reference group.

To make the link between the probability that  $X_T > X_R$  and the relative location of the two distributions more intuitive, three situations have been singled out for inspection:

<sup>23</sup> By virtue of the reference group’s mean being set to zero, target’s mean represents the difference of means between the groups.

- A)  $\mu_T = 2.00$  and  $\sigma_T = 0.75$ , with the corresponding  
 $p(X_T > X_R) = 0.945$ ;
- B)  $\mu_T = 1.00$  and  $\sigma_T = 1.00$ , with the corresponding  
 $p(X_T > X_R) = 0.760$ ;
- C)  $\mu_T = 0.25$  and  $\sigma_T = 1.50$ , with the corresponding  
 $p(X_T > X_R) = 0.555$ .

In A, the stereotype could be considered to be very accurate as it would produce mistakes when judging individuals only 5.5% of the time. In C, it is barely acceptable, with failure rate at 44.5%. What about B? With the error rate at 24% it seems quite unreliable. But before discarding it as a far-fetched overgeneralization, one should recognize that according to a popular metric proposed by Cohen (1988), this stereotype signifies a definitely large effect size. With Cohen's  $d$  equal to 1, it is the kind of effect that any social scientist would be more than happy to find in his or her study. Moreover, let us bear in mind that for "new areas of research inquiry," effect sizes of 0.2 would not be utterly uninteresting (Cohen, 1988, p.25). And it is just the effect size of a stereotypic difference depicted by C. Why should we then subject stereotypes to higher standards of truth than scientific findings?

For continuous variables,  $p(X_T > X_R)$  may be taken directly as an unambiguous measure of the stereotype's accuracy because probability that both individuals will happen to have equal values on  $X$  equals zero. Thus if we fail to observe  $X_T > X_R$ , the reverse,  $X_T < X_R$ , will be true. This is emphatically not the case with discrete variables. The smaller the number of categories on the trait variable, the more likely it is that the two individuals will be indistinguishable with regard to  $X$ . The number of  $X_T = X_R$  pairs will grow at the expense of both  $X_T > X_R$  and  $X_T < X_R$ .

Figure 2 illustrates the effect of categorization on the probabilities. Three simulations were done, each involving 100 “persons” (50 from the target group, and 50 from the reference group). Each group of 50 values constituted a representative selection from the underlying continuous distribution with standard deviation equal to 1. In terms of Cohen’s  $d$ , simulation no. 1 exemplified a moderate size effect ( $\mu_T - \mu_R = 0.5$ ), simulation no. 2 a large effect ( $\mu_T - \mu_R = 1.0$ ), and simulation no. 3 a very large one ( $\mu_T - \mu_R = 1.5$ ). In each case, 100 values from both groups were jointly ranked and grouped into 20 categories, then into 19, then 18, and so on until the final rough division into two categories: those above and those below the median. With the original continuous variables, the probabilities  $p(X_T > X_R)$  were 0.64, 0.77, and 0.87 for moderate, strong, and very strong effect sizes, respectively (see the black bullet markers in Figure 2). These probabilities decrease systematically, even though very slightly at first, with the ongoing reduction in the number of categories. Serious distortions occur when the distributions are reduced to three or, especially, two categories.<sup>24</sup>  $X_T = X_R$  probabilities are steadily on the rise, and for the moderate effect ( $\mu_T - \mu_R = 0.5$ ), we observe that splitting the continuum into two parts even makes  $X_T = X_R$  the most likely outcome. On the positive side, probabilities are not that heavily affected by reduction to five categories—the number of points on a Likert-type scale routinely applied to measure attitudes in surveys.

---

<sup>24</sup> For example,  $p(X_T > X_R) = 0.87$  associated with a very large effect on a continuous trait, dwindles to 0.61 after reducing the trait to a dichotomy—still noticeable but far from impressive.

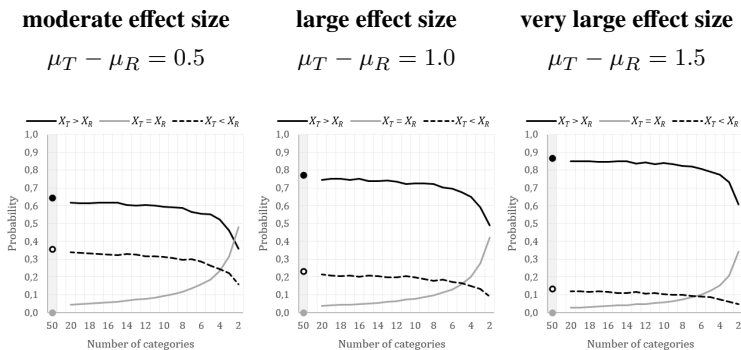


Figure 2: The effect of categorization on the observed probabilities of  $X_T > X_R$ ,  $X_T = X_R$ , and  $X_T < X_R$  for three different effect sizes (all underlying distributions are normal with  $\sigma_T = \sigma_R = 1$ ).

When analyzing the accuracy of stereotypes regarding traits measured on scales with a limited number of categories it is thus advisable to consider all three possible options:  $X_T > X_R$ ,  $X_T = X_R$ , and  $X_T < X_R$ <sup>25</sup>, as it may well turn out that  $p(X_T = X_R)$  is far from negligible.

## 8. A few empirical examples of assessing the accuracy of sex stereotypes

Both actual differences between the sexes and associated beliefs about these differences were intensively studied (Williams and Best, 1982; 1990; Costa, Terracciano and McCrae, 2001; Biernat and Deaux, 2000; Buss, 2003; Pinker, 2003). In the following, I will

<sup>25</sup> The effect of categorization cannot be held constant by using the contrast between  $p(X_T > X_R)$  and  $p(X_T < X_R)$  as a measure. The difference  $p(X_T > X_R) - p(X_T < X_R)$  tends to decrease, and the ratio  $p(X_T > X_R) / p(X_T < X_R)$  tends to increase as the number of categories becomes smaller and smaller.

examine a number of possible sex stereotypes, testing their accuracy against the survey data from Polish General Social Survey (Cichomski, Jerzyński and Zieliński, 2013) and the Study of Human Capital in Poland (Górniak, Czarnik, Turek et al., 2015; Górniak, Czarnik, Kasperek et al., 2019).<sup>26</sup> It is irrelevant here whether most people (or any people, for that matter) subscribe to these particular stereotypes, although certainly some of the inspected traits are popularly held to be gender-differentiating (e.g. body height, earnings, interest in fashion or technology).

For each trait, probabilities of a randomly selected woman being at a higher, the same, or lower level on the trait as compared to a randomly selected man were computed, using the following formulae, where  $v$  stands for a value of the trait variable<sup>27</sup>:

$$p(X_F > X_M) = \sum_{v=\min}^{\max} p(X_F = v) \cdot p(X_M < v)$$

$$p(X_F = X_M) = \sum_{v=\min}^{\max} p(X_F = v) \cdot p(X_M = v)$$

$$p(X_F < X_M) = \sum_{v=\min}^{\max} p(X_F = v) \cdot p(X_M > v)$$

Figures 3 to 6 summarize the results.

<sup>26</sup> Representative surveys provide us with proper criteria for assessing stereotypes' accuracy.

<sup>27</sup> It may be noted that a difference between the "mirror-image" probabilities, i.e.  $p(X_F > X_M) - p(X_F < X_M)$ , is equal in value to asymmetric Somer's  $d$  rank correlation coefficient (Somers, 1962), when predictions are based on subjects' sex.

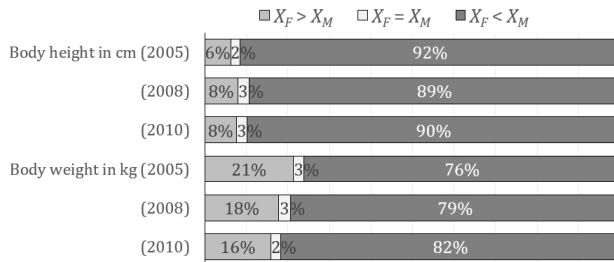


Figure 3: Probabilistic assessment of stereotypical sex differences concerning body structure.

Own computation from Polish General Social Survey data elicited by open-ended questions “What is your height and weight, approximately?”.

Both height and weight are near-continuous variables and therefore it’s very unlikely that the two randomly selected individuals will be of the same height or weight. The probabilistic differential is very large indeed, especially with regard to height. Thus the “men are taller than women” stereotype will hold true for about 90% of the randomly matched male–female pairs. It is about 12 times more likely that a man will be the taller of the two than the other way round.

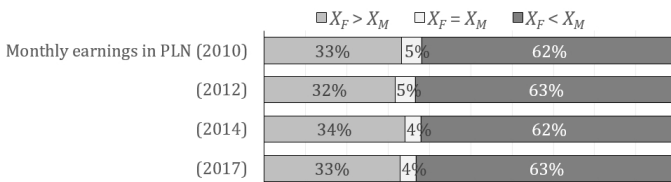


Figure 4: Probabilistic assessment of stereotypical sex differences concerning pay.

Own computation from the Study of Human Capital in Poland data elicited by open-ended questions with supplementary categorization (analysis restricted to employees with job contracts working full-time within last month, excluding self-employed).



Sex difference in average remuneration, widely publicized under the heading of “gender pay gap”, is common knowledge these days. However, the stereotypic assertion that men earn more than women is a rule with quite a lot of exceptions. In Poland, it is correct for more than 60% of the random individual comparisons, yet every third time the reverse is true. Interestingly, hardly anybody is prompted by such substantial amount of “counterevidence” to dismiss male-female wage differential as a mere stereotype.<sup>28</sup> On the contrary, the issue is widely researched, prominent in political agendas, and dealt with in antidiscrimination legislation.

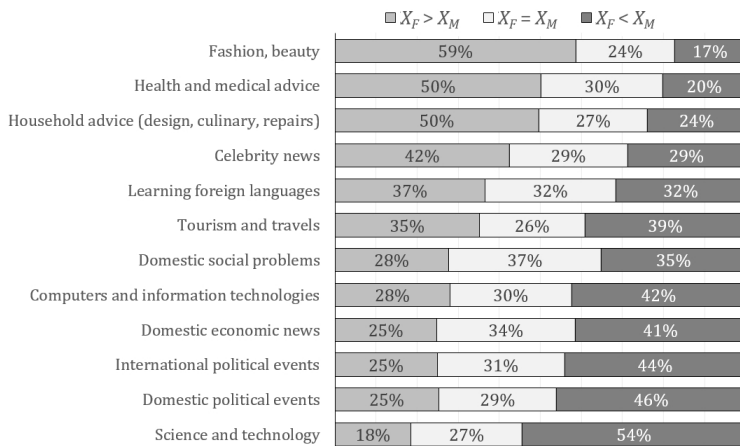


Figure 5: Probabilistic assessment of stereotypical sex differences concerning interest in the mass media content.

Own computation from Polish General Social Survey data (2002). Interest was measured on a 4-point scale: 1 not at all / 2 not much / 3 quite a bit / 4 very much.

<sup>28</sup> Similarly, Browne (1998) observes that while temperamental differences between the sexes (e.g. in competitiveness and risk-taking) are likely to be brushed aside because they “do not hold true for all individuals”, the very same argument is hardly ever used to shrug off the economic differences.

In Figure 5, people's interests are sorted by gender differences, from those most typical of women at the top to those most typical of men, at the bottom. Some of those differences are congruent with popular sex stereotypes (e.g. those concerning fashion or technology), but even then many exceptions arise. As may be expected for traits measured using categorical scales, a substantial number of male-female comparisons produce gender-equal outcomes. Still, it is unsurprising why someone seeking medical advice from a random stranger would turn for help to a woman rather than a man. As far as a man and a woman differ in their levels of medical interests and as far as the intensity of such expressed interests is proportional to actual cognizance, it is a woman who is about 2.5 times more likely to have some acquaintance with health topics. By the same token, if one seeks technological advice, it's three times more likely that of any two random strangers, one man and one woman, the man will be the one more up-to-date with matters of technology.

One can observe that stereotypical disparities in skill self-evaluations are in accord with the kinds of jobs men and women tend to perform in the gender-segregated labor market (Czarnik and Kasparek, 2015). The probabilistic skill differentials are most pronounced in the areas heavily dominated by either sex, i.e. clerical jobs associated with office and bookkeeping skills (women) and jobs in industry, construction, and transportation involving machines and technological environment (men). On the other hand, we see that "men perceive themselves to be natural managers" kind of stereotype is true by such a thin margin that it's hardly worth mentioning (see "Coordinating other persons' work").

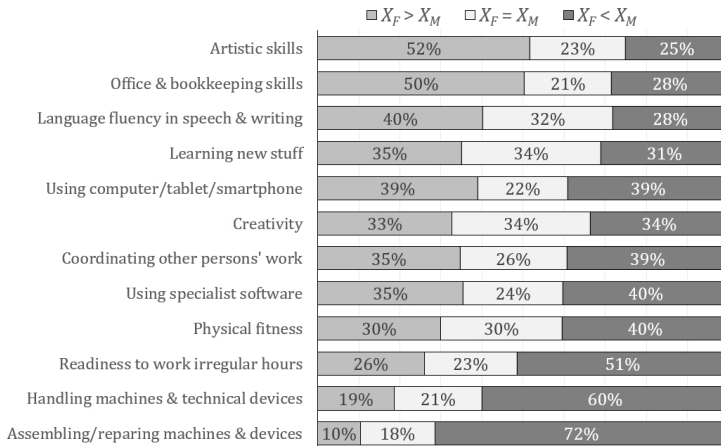


Figure 6: Probabilistic assessment of stereotypical sex differences concerning skill self-evaluation.

Own computation from the Study of Human Capital in Poland data (2017). Skill level was self-evaluated on a 5-point scale: 1 low /2 basic /3 medium /4 high /5 very high.

## 9. How good are people in assessing accuracy of stereotypes?

As shown above, stereotypic assertions in their generic formulation may vary substantially in how much truth they contain when applied at individual level and it would be very foolish of stereotype holders, namely all of us, not to recognize the fact. Fortunately, as Schneider (2005, p.199) put it, “[n]ot always, to be sure, but still most of us are aware that stereotypes are only probabilistically true and do not apply to everyone.” And even if it was true, as Pinker (2007, p.86) claims, that “[t]he image of one orb floating above another seems to come more naturally to the mind than an image of two overlapping bell curves”, people are capable of implicitly processing this kind of

information to arrive at probabilistic statements. They may commit a number of mistakes in the process, as observed by Whitley and Kite (2010, p.100):

Research participants might be fairly accurate, for example, in their estimates of what percentage of Asians are mathematical, but they might be inaccurate in their estimate of the variability of this characteristic. If perceivers are accurate on one measure, but not the other, does their belief have a kernel of truth? This question is difficult to answer.

Yet this question *can* be answered by applying the probabilistic approach. It is even more to the point for the very reason that it is our probabilistic beliefs that inform our decisions concerning individuals. A typical question you would ask yourself when pressed to choose, with certain aim in mind, one from two or more persons would not be “what is the difference of means between the groups to which these persons belong?”, and even less “how large is the dispersion of the relevant characteristic?”. It would be “which one of them is most likely to give me what I need?”. This probabilistic perception can easily be measured by asking people to imagine a familiar situation and make the assessment of comparative probabilities. By means of an illustration, consider how subjects could assess the accuracy of the stereotypic “men are taller than women” assertion.

Imagine that you enter an elevator in a large shopping mall and two random persons—one man and one woman—unacquainted to you and to each other,<sup>29</sup> enter after you. How likely it is that: (1) the man will be the taller of

---

<sup>29</sup> It is important to exclude the possibility that a man and a woman are a couple, so that subjects’ judgment about general differences would not be contaminated by their beliefs about relative height of mated individuals.

the two, (2) the woman will be the taller of the two, (3) they will be of equal height (with precision of, say, 1 cm). Make sure that probabilities add to 100%.<sup>30</sup>

A very welcome aspect of this approach is that it explicitly tackles the question of whether subjects' beliefs about certain group(s) of people are of a rigid, all-or-none, kind. This should provide solid empirical evidence to inform the dispute over sensibility of making stereotypes untrue by definition. In all likelihood, for many traits it will turn out that actually very few people incorrectly perceive the stereotyped group(s) as completely homogeneous.

Subjective probabilities elicited in this way could then be compared to the factual probabilities calculated from the source data, e.g. representative survey. For example, in the particular case of the male-female height comparison, my educated guess—informed by a number of informal trials—would be that people tend to *underestimate* the accuracy of the height stereotype.

To be sure, we still should be cautious not to make too much of such evidence. It is entirely plausible that even if people make nuanced probabilistic judgments when explicitly prompted to make them, they can still—especially in the absence of such prompts—behave automatically and unreflectively fall prey to the rigid all-or-nothing kind of thinking when making their everyday decisions. I think that an especially worthwhile avenue for further research is to investigate the relationship between the perceived accuracy of a given stereotype and perceiver's openness to, or even active seeking of, individuating information when making decisions.

---

<sup>30</sup> It is a subject for some experimentation to find out how this task could be facilitated by a particular method of registering the probabilities (e.g. by a visual representation of probabilities on screen).

## 10. Conclusion

Since the introduction of the concept into academic realm, a topic of stereotype accuracy was and remains controversial. As negative stereotypes have been linked to prejudice and discrimination, it seemed justified to view stereotypic beliefs as anything but accurate depictions of reality. However, this wholesale dismissal, insofar as it extends to all perceptions of group differences, has serious disadvantages. First, the larger the actual differences between the groups, the more obvious it is that the stereotype is (to some extent) accurate, and all statements to the contrary sound deceitful. In effect, all efforts at eradicating demonstrably false stereotypes may be hindered, as the criticism of stereotypes will lose its *prima facie* credibility. Second, outright denial of group differences where they do exist and contribute to the perceived inequality between groups severely restricts the repertoire of actions that could be taken to address the problem.

Even if many group differences turn out to be real rather than imaginary, bigots—to reiterate Kenny (1994, p.212)—“should not take any comfort in these conclusions.”<sup>31</sup> Such differences leave space—in some instances a whole lot of space—for exceptions. They also evolve: they may grow bigger, but equally likely they may grow smaller, vanish or even be reversed. In any case, to minimize the stereotype-based discrimination, it is always advisable to advocate the individualistic approach, i.e. to explain why the best decisions are the ones based—whenever possible and to whatever extent feasible—on person’s individual characteristics rather than on the average characteristics of the groups individuals belong to.

---

<sup>31</sup> Bigots should also be well advised to recognize that they themselves are a strongly stereotyped group, and these stereotypes may turn out to be not so far from the truth, either.

Testing stereotype's accuracy in probabilistic terms has a direct advantage of bridging the gap between collective and individual levels of analysis, both of which are explicitly involved in making probabilistic comparisons. This, to satisfy Jussim (2012, p.309), should bring "greater conceptual clarity [...] to understanding stereotype accuracy." It could also be a step towards meeting concerns voiced by Stangor (1995, p.278) that in order to find "an adequate remedy for negative intergroup behavior [...] We need to know both how big existing group differences are and how to communicate the extent of those differences to people." It seems that probabilistic approach is quite a promising avenue in this regard by explicitly bringing up counterexamples to the stereotype under question.

In the end, no matter how much—or how little—truth any particular stereotypes may contain, "what matters is [...] the gullibility with which we employ them". And we would be wise indeed to take this advice from Walter Lippmann, which is no less prudent today than it was a hundred years ago:

if our philosophy tells us that each man is only a small part of the world, that his intelligence catches at best only phases and aspects in a coarse net of ideas, then, when we use our stereotypes, we tend to know that they are only stereotypes, to hold them lightly, to modify them gladly (Lippmann, 1998, pp.90–91).

**Acknowledgment** Insightful comments by two anonymous reviewers and a mathematical consultation from Dr. Paweł Hołobut (IPPT PAN) are gratefully acknowledged.

## Bibliography

- Allport, G.W., 1954. *The Nature of Prejudice*. Reading, Mass.: Addison-Wesley Pub. Co.
- Aronson, E., 2011. *The Social Animal*. New York: Worth Publishers.
- Arrow, K.J., 1971. *The theory of discrimination*. Working Paper 30A. Princeton, N.J.: Princeton University.
- Ashmore, R. and Del Boca, F., 1981. Conceptual approaches to stereotypes and stereotyping. In: D.L. Hamilton, ed. *Cognitive Processes in Stereotyping and Intergroup Behavior*. London and New York: Psychology Press, Taylor & Francis Group, pp.1–36.
- Banaji, M., 2001. Stereotypes, Social Psychology of. In: N.J. Smelser and P.B. Baltes, eds. *International Encyclopedia of the Social & Behavioral Sciences*. Amsterdam; New York: Elsevier, pp.15100–15104.
- Bar-Tal, D., 1989. Delegitimization: the extreme case of stereotyping and prejudice. In: D. Bar-Tal, C.F. Graumann, A.W. Kruglanski and W. Stroebe, eds. *Stereotyping and Prejudice: Changing Conceptions* [Online], *Springer Series in Social Psychology*. New York, NY: Springer, pp.169–182. Available at: [https://doi.org/10.1007/978-1-4612-3582-8\\_8](https://doi.org/10.1007/978-1-4612-3582-8_8) [visited on 17 July 2020].
- Beeghly, E.M., 2014. *Seeing Difference: The Ethics and Epistemology of Stereotyping* [Online]. Ph.D. dissertation. Berkeley: University of California. Available at: <[https://digitalassets.lib.berkeley.edu/etd/ucb/text/Beeghly\\_berkeley\\_0028E\\_14475.pdf](https://digitalassets.lib.berkeley.edu/etd/ucb/text/Beeghly_berkeley_0028E_14475.pdf)> [visited on 17 July 2020].
- Benbow, C.P. and Lubinski, D., 1993. Psychological profiles of the mathematically talented: some sex differences and evidence supporting their biological basis. *Ciba Foundation Symposium* [Online]. Vol. 178, pp.44–59, discussion pp.60–66. Available at: <https://doi.org/10.1002/9780470514498.ch4> [visited on 17 July 2020].
- Biernat, M. and Deaux, K., 2000. Sex differences. In: E.F. Borgatta and R.J.V. Montgomery, eds. *Encyclopedia of Sociology*. 2nd ed. Vol. 4. New York: Macmillan Reference, pp.2529–2536.
- Browne, K., 1998. *Divided Labours: An Evolutionary View of Women at Work, Darwinism Today*. London: Weidenfeld & Nicolson.



- Buss, D.M., 2003. *The Evolution of Desire: Strategies of Human Mating*. Rev. ed. New York: Basic Books.
- Cichomski, B., Jerzyński, T. and Zieliński, M., 2013. *Polskie Generalne Sondáže Społeczne: struktura skumulowanych wyników badań 1992-2010*. Warszawa: Instytut Studiów Społecznych, Uniwersytet Warszawski.
- Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale: Lawrence Erlbaum Associates Publishers.
- Costa, P.T., Terracciano, A. and McCrae, R.R., 2001. Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology* [Online], 81(2), pp.322–331. Available at: <https://doi.org/10.1037/0022-3514.81.2.322> [visited on 17 July 2020].
- Cronbach, L.J., 1955. Processes affecting scores on "understanding of others" and "assumed similarity". *Psychological Bulletin* [Online], 52(3), pp.177–193. Available at: <https://doi.org/10.1037/h0044919> [visited on 17 July 2020].
- Czarnik, S. and Kasperek, K., 2015. Gender segregation in the Polish labour market. *The Hidden Human Capital: Additional Capacities of the Polish Labour Market: Key Results of the Fifth Round of the BKL Study in 2014* [Online]. Warszawa - Kraków: Polish Agency for Enterprise Development, pp.37–56. Available at: <<https://ruj.uj.edu.pl/xmlui/handle/item/62666>> [visited on 17 July 2020].
- English, H.B. and English, A.C., 1958. *A comprehensive dictionary of psychological and psychoanalytical terms: a guide to usage*. London: Longmans, Green, and Co.
- Fisher, A.J., Medaglia, J.D. and Jeronimus, B.F., 2018. Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences* [Online], 115(27), E6106–E6115. Available at: <https://doi.org/10.1073/pnas.1711978115> [visited on 17 July 2020].
- Górniak, J., Czarnik, S., Kasperek, K. et al., 2019. *Bilans Kapitału Ludzkiego: badanie ludności: skumulowana baza danych 2017-2018*.

- Kraków; Warszawa: Centrum Ewaluacji i Analiz Polityk Publicznych Uniwersytetu Jagiellońskiego, Polska Agencja Rozwoju Przedsiębiorczości.
- Górniak, J., Czarnik, S., Turek, K. et al., 2015. *Bilans Kapitału Ludzkiego: badanie ludności: skumulowana baza danych 2010-2014*. Kraków; Warszawa: Centrum Ewaluacji i Analiz Polityk Publicznych Uniwersytetu Jagiellońskiego, Polska Agencja Rozwoju Przedsiębiorczości.
- Jones, E.E. and Colman, A.M., 2001. Stereotypes. In: A. Kuper and J. Kuper, eds. *The Social Science Encyclopedia*. 2nd ed. London; New York: Routledge, pp.1449–1450.
- Judd, C.M. and Park, B., 1993. Definition and assessment of accuracy in social stereotypes. *Psychological Review* [Online], 100(1), pp.109–128. Available at: <https://doi.org/10.1037/0033-295X.100.1.109> [visited on 17 July 2020].
- Jussim, L., 2012. *Social Perception and Social Reality: Why Accuracy Dominates Bias and Self-Fulfilling Prophecy*. Oxford [etc.]: Oxford University Press.
- Jussim, L., Cain, T.R. et al., 2009. The unbearable accuracy of stereotypes. *Handbook of Prejudice, Stereotyping, and Discrimination*. New York, NY: Psychology Press, Taylor & Francis Group, pp.199–227.
- Jussim, L., Crawford, J.T., Anglin, S.M. et al., 2016. Stereotype accuracy: One of the largest and most replicable effects in all of social psychology. *Handbook of Prejudice, Stereotyping, and Discrimination*. 2nd ed. New York, NY: Psychology Press, Taylor & Francis Group, pp.31–63.
- Jussim, L., Crawford, J.T. and Rubinstein, R.S., 2015. Stereotype (in)accuracy in perceptions of groups and individuals. *Current Directions in Psychological Science* [Online], 24(6), pp.490–497. Available at: <https://doi.org/10.1177/0963721415605257> [visited on 17 July 2020].
- Kenny, D.A., 1994. *Interpersonal Perception: A Social Relations Analysis, Distinguished Contributions in Psychology*. New York; London: The Guilford Press.
- Klineberg, O., 1951. The scientific study of national stereotypes. *International Social Science Bulletin*, 3(3), pp.505–514.

- Koleser, J., 2008. Stereotyping. In: V.N. Parrillo, ed. *Encyclopedia of Social Problems*. Los Angeles; London; New Delhi; Singapore: Sage, pp.900–902.
- Lee, T. and Fiske, S., 2008. Stereotypes. In: W.A.J. Darity, ed. *International encyclopedia of the social sciences*. 2nd ed. Vol. 8. Detroit, Mich.: Macmillan Reference USA, pp.136–139.
- Lee, Y.-T., Jussim, L.J. and McCauley, C.R., 1995. *Stereotype Accuracy: Toward Appreciating Group Differences* [Online]. Washington, DC: American Psychological Association. Available at: <<http://content.apa.org/books/2003-88371-000>> [visited on 17 July 2020].
- Leslie, S.-J. and Lerner, A., 2016. Generic Generalizations. In: E.N. Zalta, ed. *The Stanford Encyclopedia of Philosophy* [Online]. Winter 2016. Stanford: Metaphysics Research Lab, Stanford University. Available at: <<https://plato.stanford.edu/archives/win2016/entries/generics/>> [visited on 17 July 2020].
- Lippmann, W., 1998. *Public Opinion*. 2nd ed. New Brunswick; London: Transaction Publishers.
- Macrae, C.N. and Quadflieg, S., 2010. Perceiving people. *Handbook of Social Psychology, Vol. 1, 5th Ed* [Online]. 5th ed. Vol. 1. Hoboken, NJ: John Wiley & Sons, Inc., pp.428–463. Available at: <https://doi.org/10.1002/9780470561119.socpsy001012> [visited on 17 July 2020].
- McCauley, C., Stitt, C.L. and Segal, M., 1980. Stereotyping: From prejudice to prediction. *Psychological Bulletin* [Online], 87(1), pp.195–208. Available at: <https://doi.org/10.1037/0033-2909.87.1.195> [visited on 17 July 2020].
- Merton, R.K., 1948. The self-fulfilling prophecy. *The Antioch Review* [Online], 8(2), p.193. Available at: <https://doi.org/10.2307/4609267> [visited on 17 July 2020].
- Miller, D.T. and Turnbull, W., 1986. Expectancies and interpersonal processes. *Annual Review of Psychology* [Online], 37(1), pp.233–256. Available at: <https://doi.org/10.1146/annurev.ps.37.020186.001313> [visited on 17 July 2020].

- Phelps, E.S., 1972. The statistical theory of racism and sexism. *The American Economic Review* [Online], 62(4), pp.659–661. Available at: <<https://www.jstor.org/stable/1806107>> [visited on 17 July 2020].
- Pickering, M., 2011. Stereotyping and stereotypes. In: G. Ritzer and J. Ryan, eds. *The Concise Encyclopedia of Sociology*. Wiley-Blackwell, pp.616–617.
- Pinker, S., 2003. *The Blank Slate: The Modern Denial of Human Nature*. London: Penguin Books.
- Pinker, S., 2007. *The Stuff of Thought: Language as a Window Into Human Nature*. New York: Viking.
- Schauer, F.F., 2003. *Profiles, Probabilities, and Stereotypes*. Cambridge, Mass.; London: Belknap Press of Harvard University Press.
- Schneider, D.J., 2005. *The Psychology of Stereotyping*. New York; London: Guilford.
- Somers, R.H., 1962. A new asymmetric measure of association for ordinal variables. *American Sociological Review* [Online], 27(6), pp.799–811. Available at: <https://doi.org/10.2307/2090408> [visited on 17 July 2020].
- Sowell, T., 1995. *The Vision of the Anointed: Self-Congratulation as a Basis for Social Policy*. New York: Basic Books.
- Stangor, C., 1995. Content and application inaccuracy in social stereotyping. In: Y.-T. Lee, L.J. Jussim and C.R. McCauley, eds. *Stereotype Accuracy: Toward Appreciating Group Differences*. [Online]. Washington: American Psychological Association, pp.275–292. Available at: <https://doi.org/10.1037/10495-011> [visited on 17 July 2020].
- Stroebe, W. and Insko, C.A., 1989. Stereotype, prejudice, and discrimination: changing conceptions in theory and research. In: D. Bar-Tal, C.F. Graumann, A.W. Kruglanski and W. Stroebe, eds. *Stereotyping and Prejudice: Changing Conceptions* [Online], *Springer Series in Social Psychology*. New York, NY: Springer, pp.3–34. Available at: [https://doi.org/10.1007/978-1-4612-3582-8\\_1](https://doi.org/10.1007/978-1-4612-3582-8_1) [visited on 17 July 2020].

- Swim, J.K., 1994. Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology* [Online], 66(1), pp.21–36. Available at: <https://doi.org/10.1037/0022-3514.66.1.21> [visited on 17 July 2020].
- USHMM (United States Holocaust Memorial Museum), n.d. *Defining the Enemy*. Available at: <https://encyclopedia.ushmm.org/content/en/article/defining-the-enemy> [visited on 17 July 2020].
- Werth, N., 2011. *Dekulakisation as mass violence*. Available at: <http://dekulakisation-mass-violence.html> [visited on 17 July 2020].
- Whitley, B.E. and Kite, M.E., 2010. *The Psychology of Prejudice and Discrimination*. Belmont, CA: Wadsworth Cengage Learning.
- Wikipedia, 2020. *Stereotype*. Available at: <https://en.wikipedia.org/w/index.php?title=Stereotype&oldid=967765670> [visited on 17 July 2020].
- Williams, J.E. and Best, D.L., 1982. *Measuring Sex Stereotypes: A Thirty-Nation Study*. Beverly Hills; London; New Delhi: Sage Publications.
- Williams, J.E. and Best, D.L., 1990. *Sex and Psyche: Gender and Self Viewed Cross-Culturally, Cross-Cultural Research and Methodology Series 13*. Newbury Park, CA; London; New Delhi: Sage Publications.