Citation: Madon, S., Jussim, G., Nofziger, H. Salib, E., R., Willard, J., Scherrr, K. C. (in press). The accumulation of stereotype-based self-fulfilling prophecies. *Journal of Personality and Social Psychology.*

The Accumulation of Stereotype-Based Self-Fulfilling Prophecies

Stephanie Madon
Iowa State University

Lee Jussim
Rutgers University

Max Guyll
Iowa State University

Heather Nofziger Elizabeth Salib
Rutgers University

Jennifer Willard
Kennesaw State University

Kyle C. Scherr
Central Michigan University

Corresponding author: Stephanie Madon, madon@iastate.edu.

Correspondence concerning this article should be addressed to Stephanie Madon at W112 Lagomarcino Hall, 901 Stange Road, Psychology Department, Iowa State University, Ames, IA 50011. E-mail: madon@iastate.edu.

Abstract

A recurring theme in the psychological literature is that the self-fulfilling effect of stereotypes can accumulate across perceivers. This article provides the first empirical support for this long-standing hypothesis. In three experiments (*N*s = 123 – 241), targets more strongly confirmed a stereotype as the number of perceivers who held stereotypic expectations about them increased. A fourth experiment (*N* = 121) showed that new perceivers judged targets according to the stereotypic behaviors they had previously been channeled to adopt, an effect that even occurred among perceivers who were privy to the fact that targets' behavior had been shaped by the actions of others. The authors discuss ways in which these effects may contribute to group inequalities.

Keywords: self-fulfilling prophecy, behavioral confirmation, stereotypes, accumulation

The Accumulation of Stereotype-Based Self-Fulfilling Prophecies

New Look in Perception of the 1940s and 1950s initiated a revolution in approaches to perception within social psychology. Departing from the prevailing view that perception is veridical, New Look in Perception research promoted the idea that perception is influenced by the goals, needs, and motives of perceivers. An influential perspective that emerged from this movement was a weak form of social constructionism. According to this perspective, social beliefs can alter reality and shape behavior. The self-fulfilling prophecy is central to this perspective because it involves a perceiver's false expectation about a target initiating a sequence of events that causes the target to exhibit expectancy-consistent behavior, thereby making the initially false expectation true. This research tested a core tenet of social constructionism within social psychology – the idea that self-fulfilling prophecy effects can accumulate across perceivers. Moreover, it tested this hypothesis with respect to stereotypes, which psychological theory proposes contribute to group inequalities through their cumulative self-fulfilling effects.

Self-Fulfilling Prophecies and their Cumulative Effects

The idea that false expectations can lead to their own fulfillment originated in the writings of Merton (1948). Merton proposed that the self-fulfilling prophecy was a powerful process capable of producing profound social problems including war, economic downturns, academic underachievement, and racial disparities in employment and wealth. Research bearing on Merton's analysis clearly supported the existence of self-fulfilling prophecies, but not the idea that self-fulfilling prophecy effects are powerful. Both experimental and naturalistic research have converged on the conclusion that perceivers' false expectations have only modest self-fulfilling effects on the behavior of targets (Jussim, 2012; Rosenthal, 1994, 2003). However, these modest effects should not be interpreted to mean that self-fulfilling prophecies can never be

powerful. Even small self-fulfilling prophecy effects can become powerful if they accumulate across perceivers (e.g., Jussim, Eccles, & Madon, 1996; Klein & Snyder, 2003; Madon, Guyll, Spoth, & Willard, 2004; Merton, 1948).

The potential for self-fulfilling prophecy effects to accumulate across perceivers represents a central theme within social psychology (Ross, Lepper, & Ward, 2010). Yet, only one study has empirically supported the hypothesized effect. In the context of a longitudinal study involving parents and their adolescent children, Madon et al. (2004) found that adolescents drank the greatest amount of alcohol when mothers and fathers both held negative expectations about their future alcohol use. However, because Madon et al. demonstrated this effect with correlational data, they could not rule out predictive accuracy as an alternative explanation of the findings. Put differently, their data could not exclude the possibility that parents' negative expectations were accurate from the outset, in which case they could not have been self-fulfilling, a limitation that characterizes all correlational self-fulfilling prophecy research.

The only way to eliminate predictive accuracy as an alternative to a self-fulfilling prophecy interpretation is to experimentally manipulate perceivers' expectations. Accordingly, there is a need to examine the accumulation of self-fulfilling prophecy effects across perceivers in tightly controlled laboratory experiments. It is especially important to examine this process with respect to stereotypes, which for 70 years have been hypothesized to contribute to group inequalities via their cumulative self-fulfilling effects (Jussim et al., 1996; Klein & Snyder, 2003; Merton, 1948; Word, Zanna, & Cooper, 1974).

There is good evidence that stereotypes can have self-fulfilling effects on targets' behavior (Snyder, Tanke, & Berscheid, 1977; Word et al., 1974). However, research demonstrating this effect has focused exclusively on dyadic relations involving one perceiver and

one target. Although this focus is often warranted, it may underestimate the power of stereotypes because it does not account for the possibility that their self-fulfilling effects may accumulate across multiple perceivers. Specifically, because stereotypes can be consensual, different perceivers may hold similar expectations about members of stereotyped groups (Madon et al., 2001). To the extent that these expectations are false for a particular target, each perceiver may have a self-fulfilling effect that combines with the self-fulfilling effects of other perceivers to ultimately cause a target to confirm the stereotype more strongly than would have been the case had only one perceiver held the false stereotypic expectation.

Situational Affordances

One mechanism through which stereotypes may have cumulative self-fulfilling effects is situational affordances. Generally speaking, affordances are the properties of a stimulus that encourage a particular behavioral response (Gibson, 1979). In the words of Koffka (1935), for example, “a fruit says, ‘Eat me’; water says, ‘Drink me’..." (p. 7). Although typically applied to the properties of objects, perceivers’ treatment of targets can construct situations with opportunities or constraints that may likewise be viewed as affordances. If these opportunities and constraints channel a target to behaviorally confirm a false expectation, then the self-fulfilling prophecy occurs because the situation afforded such behavior.

Classic considerations of the self-fulfilling prophecy highlight situational affordances as a key mechanism through which stereotypes can become self-fulfilling. For example, Merton (1948) argued that the early 20th century practice of barring Blacks from labor unions on grounds that they were strike breakers encouraged Black laborers to cross picket lines by restricting their job opportunities. Similarly, Rosenthal & Jacobson (1968) and Rist (1970) hypothesized that negative expectations caused teachers to limit the educational opportunities available to

disadvantaged students, thereby undermining those students' academic achievement. These examples illustrate how perceivers' treatment of targets can construct situational affordances that encourage targets to behaviorally confirm a stereotype.

The potential for a stereotype's self-fulfilling effect to accumulate across perceivers arises when multiple perceivers provide similar situational affordances to the same target. For example, consider a scenario in which two educators independently provide a female student with a situational affordance that encourages confirmation of sex stereotypes: a math teacher inappropriately tracks her into a low-ability math class, and a guidance counselor encourages her to take home economics or typing as the required elective, never suggesting alternatives such as computer programming or woodshop. Although these situational affordances do not prevent the student from excelling in math or enrolling in a male-dominated elective, they will tend to channel her in the direction of confirming sex stereotypes. Moreover, because two educators each constructed the situation in this way, the overall effect on the student stands to be greater than if only one educator had done so. This is because multiple perceivers who share, and independently act upon, a false expectation generate multiple vectors of influence that can combine to shape the totality of the situation faced by a target in a way that more strongly affords behavioral confirmation of a stereotype than is the case with only one perceiver.

Research Overview

The primary objective of the present research was to test the hypothesis that the self-fulfilling effects of stereotypes can accumulate across perceivers. Experiment 1 tested this hypothesis with respect to the overweight stereotype, whereas Experiments 2 and 3 tested this hypothesis with respect to sex stereotypes. Consistent with classic considerations of the self-fulfilling prophecy, the experiments focused on situational affordances as the underlying

mechanism of the hypothesized accumulation effect. Experiment 4 tested whether targets' confirmatory behavior – behavior that was caused by a stereotype's prior cumulative self-fulfilling effect – influences new perceivers' judgements of them. This issue is important because it addresses the possibility that once the cumulative self-fulfilling effect of a stereotype has been set in motion, it can contribute to a cycle whereby prior self-fulfilling prophecy effects lead new perceivers to develop false expectations, which may themselves become self-fulfilling.

Analytic strategy. Although the methods of the experiments differed, the same analytic strategy was used to test the accumulation hypothesis. First, the analyses tested whether perceivers developed stereotypic expectations about a target. Second, the analyses tested whether perceivers' stereotypic expectations led them to provide targets with situational affordances that encouraged a stereotypic response. Third, the analyses examined whether the number of perceivers who held a stereotypic expectation about a target influenced how strongly the target confirmed the stereotype. Fourth, a series of planned contrasts tested for dyadic self-fulfilling prophecy effects involving one perceiver and one target, and then for the hypothesized accumulation effect which, in this research, involved two perceivers and one target. Finally, the analyses explored whether the accumulation effects reflected *concurrent* or *synergistic accumulation*. Concurrent accumulation occurs when multiple perceivers each have a unique, additive self-fulfilling effect on a target's behavior (Jussim et al., 1996). Synergistic accumulation occurs when the self-fulfilling effects of multiple perceivers are stronger in combination than their additive effects would suggest (Madon et al., 2004).

Type I error. Bonferroni corrections and HSD Tukey contrasts controlled for Type I error when there were multiple comparisons, any one of which would support a predicted effect (e.g., testing whether perceivers developed stereotypic expectations). LSD contrasts were used

when Type 1 error was not an issue because the predicted effect required only a single significant comparison (e.g., testing whether a stereotype had a cumulative self-fulfilling effect) or required multiple significant comparisons to support a predicted effect (e.g., testing for concurrent or synergistic accumulation). Multiple analysis of variance (MANOVA) was used to examine the effect of experimental factors on multiple dependent variables of the same underlying construct.

Power and effect sizes. Meta-analyses of interpersonal self-fulfilling prophecies report an average effect size of d = .60 (Rosenthal, 1994). Because the current research focused on accumulation, we increased this effect size by 10% when performing power analyses. The results indicated that Experiments 1 and 2 each required a sample size of 93 to achieve a power of .80 for detecting the anticipated accumulation effect of d = .66, whereas Experiment 3 required a sample size of 115. When calculating the sample size of Experiment 4, we used an effect size of d = .80 on grounds that the meta-analytic effect of individuating information on person perception is large (Kunda & Thagard, 1996). Using this estimate, a power analysis indicated that Experiment 4 required a sample of 80 to achieve a power of .80. All sample sizes exceeded these minimums. Effect sizes and their confidence intervals (CI) are reported in terms of d, η^2, and η_p^2, which were calculated with scripts developed by Wuensch (2012). A 95% CI is reported for d, and a 90% CI for η^2 and η_p^2 (Steiger, 2004).

Experiment 1

Experiment 1 tested whether the self-fulfilling effect of the overweight stereotype can accumulate across perceivers. The procedures involved two phases. Phase 1 manipulated a target's weight to examine whether it influenced the situational affordances that individual perceivers provided a target. Phase 2 exposed a new group of participants (designated as targets in the triads) to the combined situational affordances that two randomly paired perceivers had

provided the target in Phase 1. Targets' behavior during the experimental session indexed the degree to which they confirmed the overweight stereotype.

Method

Participants

Undergraduates (*N* = 723) at Iowa State University participated to fulfill a course requirement, including 403 women, 319 men, and one participant who omitted a response. There were 19 African Americans, 26 Asian/Asian Americans, 626 European Americans, 5 Native Americans, 5 Indians, and 42 participants who self-described as multi-ethnic.

Experimental Design

Participants were randomly assigned to triads (*N* = 241 triads) each consisting of two perceivers and one target. Each triad was randomly assigned to one of three expectation conditions. In the *no-overweight expectation condition* (*n* = 83 triads), both perceivers in a triad believed the target was thin. In the *single-overweight expectation condition* (*n* = 76 triads), one perceiver in a triad believed the target was thin, whereas the other believed the target was heavy. In the *double-overweight expectation condition* (*n* = 82 triads), both perceivers in a triad believed the target was heavy. Although triad was the unit of analysis, participants completed the experimental procedures independently; hence, the data were not nested. The targets' weight was manipulated with a bogus profile and photograph that were randomly assigned to perceivers.

Phase 1: Materials and Measures

Profile. A handwritten profile that appeared to have been completed by the target at an earlier session reported the target's sex, age, height, weight, and personality. The profile always described the target as female, European American, 21 years old, "*between 5'4 and 5'6 feet tall"*, and as having a constellation of personality traits that was held constant. The profile

systematically varied the target's weight to be *"between 101 and 120 pounds"* (*n* = 242) or *"between 181 and 200 pounds"* (*n* = 240).

Photograph. A photograph of a heavy or thin European American woman in her early twenties accompanied each profile. For stimulus sampling purposes, two photographs depicted a heavy woman and two depicted a thin woman.

Perceiver behavior. Perceivers predicted the target's attitudes and behaviors on a variety of issues by responding to a series of questions. Included among these was a critical item that assessed the extent to which perceivers provided the target with a situational affordance that encouraged confirmation of the overweight stereotype. Specifically, perceivers were shown colored pictures of four bins that contained 2 (Bin 1), 4 (Bin 2), 20 (Bin 3), and 40 (Bin 4) pieces of candy, and asked "*Which bin should be given to the person in the next phase? Choose the bin that you think contains the approximate amount of candy the person would eat if nobody was around."* Four types of candy were used: Kisses, Butterfingers, Peanut Butter Cups, and Kit-Kats, the last three of which were all *fun-size*. Except for Bin 1, which contained one Butterfinger and one Kit-Kat, equal numbers of each type of candy were included in each bin.

Although the amount of candy in the bins was arbitrary, the spread was by design. We did not want perceivers to have the option of selecting an intermediate amount of candy because such an option would have encouraged them to hedge. We wanted to compel perceivers to make a clear choice: choose a little or choose a lot. We included some variation on each side so that perceivers did not feel overly restricted in their choices. To reduce suspicion, the critical item was embedded among fillers each accompanied by a colored picture (i.e., how much the target liked popular TV shows; target's likelihood of using an electronic voting machine and taking an on-line college course; how favorably the target viewed DNA testing).

Manipulation checks. To assess whether perceivers noticed the target's weight, they reported the target's weight category as indicated on the profile. This question was embedded among fillers that instructed perceivers to report the target's sex and height which were also indicated on the profile. To assess whether the target's weight activated perceivers' stereotypes, perceivers judged the target with respect to five traits, including two that are stereotypic of heavy people (*willpower, self-control*) and three that are not strongly associated with weight (*outgoing, intelligent, religious*; Crandall, 1994; Puhl & Brownell, 2001). Perceivers made these trait judgments on 7-point scales with endpoints 1 (*not at all*) and 7 (*very much*).

Suspicion check for perceivers. Perceivers reported their beliefs about the experiment's purpose, research questions under investigation, and any prior knowledge they had of the study.

Phase 2: Materials and Measures

Candy. Each target received one bowl of candy that included the exact amount and type of candy that the two perceivers in the target's triad had selected for the target in Phase 1. For example, if one perceiver selected Bin 1 (i.e., one Butterfinger and one Kit-Kat) and the other perceiver selected Bin 2 (i.e., one Kiss, one Butterfinger, one Peanut Butter Cup, and one Kit-Kat), then the target in this triad would have received one bowl of candy that included one Kiss, two Butterfingers, one Peanut Butter Cup, and two Kit-Kats. Thus, each perceiver independently selected a bin of candy for the target in Phase 1, and the candy they selected was combined and given to a target in Phase 2. This procedure created a situation that is analogous to real-world circumstances in which the independent actions of multiple perceivers (e.g., teachers, employers, parents) can combine to affect how strongly the situation faced by a target (e.g., student, employee, child) encourages behavioral confirmation of a stereotype.

Target behavior. The amount of candy targets took from the bowl provided an explicit

behavior with which to determine how strongly they confirmed the overweight stereotype.

Filler questions and props. To support the cover story used in Phase 2 – i.e., that the study was designed to examine the relationship between personality and taste preferences – all targets completed filler questions that assessed their personality traits and expectations about the candy they expected to eat. In addition, several props encouraged targets to take candy home. A printed sign read "*Feel free to take as much candy home with you as you like. We have plenty"*. There was a large stack of brown paper lunch bags provided for the purpose of carrying the candy. A garbage can filled with candy wrappers was placed on the floor near targets. At the end of the session, a computer message encouraged targets to take candy home.

Suspicion check for targets. Targets described their beliefs about the experiment's purpose and reported any prior knowledge they had of the study.

Procedures

Phase 1. Perceivers were run in group sessions, but provided independent responses. After obtaining informed consent, the experimenter described the study as examining how accurately people can predict other people's attitudes and behaviors from their personalities. Perceivers then received a packet that included the target's profile and photograph, plus a survey. Although perceivers believed the target was another participant in the study and unique to them, multiple copies of the four different photographs were distributed at each experimental session, but ordered in a way that prevented adjacent perceivers from receiving the same one.

To substantiate the target's authenticity, the experimenter asked perceivers to indicate whether they knew the person assigned to them. A stooge planted in the group publicly stated knowing the person in her or his packet. In response, the experimenter gave the stooge a new packet and asked if s/he knew this second person. The stooge always indicated that s/he did not,

at which point the experimenter again asked if there were others who knew the person in their respective packets. When it was confirmed that no one did, the experimenter explained that the individuals in the packets had completed a profile and provided a photograph at an earlier session, and would return for another session to have their attitudes and behaviors assessed.

The experimenter then directed perceivers to use the target's profile to complete the accompanying survey, which instructed them to predict the target's attitudes and behaviors on the same issues they believed the target would provide information about at the later session. Perceivers believed that their predictions would be compared to the target's actual responses. Perceivers retained the profile and photograph while making their predictions. Afterward, participants completed the manipulation and suspicion checks and provided demographic information. Debriefing followed. The stooge completed all materials along with perceivers.

Phase 2. After the data from Phase 1 had been collected, the perceivers were randomly paired to create the expectation conditions. The candy selected by each pair of perceivers was combined and given to a third participant who was randomly assigned to be the target in the triad. After providing informed consent, the experimenter escorted each target to a private room equipped with a computer, bowl of candy, and props. The experimenter explained that the study was designed to examine how personality relates to taste preferences, and informed targets that they would perform a taste test of candy immediately after answering survey questions. The target was further informed that the experimenter would wait for the target to complete the experiment in a separate lab, and that the target should come to that location at the end of the session to receive research credit. This procedure provided targets with complete privacy, thereby reducing any inhibitions they might have had taking candy. After completing the survey, which included filler questions and suspicion checks, the computer informed targets that they

would not engage in a taste test of the candy after all, but were free to take as much as they wanted. Targets then met the experimenter as instructed, and were debriefed. Upon their departure, the experimenter returned to the testing room to record the amount of candy taken.

Results

Preliminary Analyses

Descriptive statistics. Supplementary Tables 1 and 2 report descriptive statistics for the variables in Phase 1 and Phase 2, respectively.

Manipulation checks. There were 480 perceivers (> 99%) who correctly reported the target's weight and 2 (< 1%) who omitted a response. Independent sample *t*-tests, with a Bonferroni correction criterion of $p \leq .01$, indicated that perceivers judged the target as having significantly less willpower and self-control when they believed the target was heavy versus thin, ts ≥ 3.98, ps $< .001$; perceivers' judgments of the target's outgoingness, intelligence and religiosity did not differ significantly as a function of the target's weight, ts ≤ 1.46, ps $\geq .146$. These results show that perceivers developed stereotypic expectations about the target. Table 1 reports the means, standard deviations, and effect sizes associated with these results.

Suspicion checks. There were five suspicious perceivers and two suspicious targets, no two from the same triad. Excluding their data slightly increased the expectation's effect, but did not meaningfully alter the results. No participant's data were excluded from the analyses.

Main Analyses

The manipulation checks demonstrated that perceivers developed stereotypic expectations about the target. Therefore, the main analyses tested for a series of effects that are relevant to the overweight stereotype's cumulative self-fulfilling effect.

Situational affordances. First, the analyses tested whether perceivers' stereotypic

expectations caused them to provide the target with a situational affordance that encouraged confirmation of the overweight stereotype. A *Mann Whitney U* examined whether the target's weight (heavy vs. thin) influenced the bin of candy that individual perceivers selected for the target in Phase 1. The results showed that perceivers chose bins containing significantly more candy when they believed the target was heavy (*Mdn* = Bin 3, 20 pieces of candy) versus thin (*Mdn* = Bin 2, 4 pieces of candy), $U = 12279$, $p < .001$. Thus, individual perceivers provided the target with a situational affordance that more strongly encouraged confirmation of the overweight stereotype when they believed the target was heavy versus thin.

In addition, a one-way ANOVA tested whether the number of perceivers who believed the target was heavy influenced the amount of candy targets received in Phase 2. The independent variable was the expectation (no-overweight vs. single-overweight vs. double-overweight). The dependent variable was the amount of candy given to targets in Phase 2. The results showed that targets received the least candy in the no-overweight expectation condition ($M = 17$, $SD = 13.14$), an intermediate amount of candy in the single-overweight expectation condition ($M = 27$, SD 14.82), and the most candy in the double-overweight expectation condition ($M = 42$, $SD = 15.74$), $F(2, 238) = 59.01$, $p < .001$, $\eta^2 = .33$, 90% CI [0.25, 0.40]. Three Tukey HSD contrasts indicated that these amounts all differed significantly from one another, ts$(238) \geq 3.98$, ps $< .001$, ds $\geq .63$. These results demonstrate that the situation constructed by two, independent perceivers afforded targets more opportunity to confirm the overweight stereotype than did the situation constructed by individual perceivers singly.

Target behavior. Second, the analyses tested whether the number of perceivers who held a stereotypic expectation about the target influenced how strongly the target confirmed the stereotype. A one-way ANOVA indicated differences in the effect of the expectation (no-

overweight vs. single-overweight vs. double-overweight) on the amount of candy taken by targets in Phase 2, $F(2, 238) = 5.32$, $p = .005$, $\eta^2 = .043$, 90% CI [0.01, 0.09]. A mediational path model showed that the amount of candy targets received explained 86% of the expectation's effect, $b = .744$; $SE = .228$, $p < .001$, suggesting nearly full mediation.

Self-fulfilling prophecy effect. Third, an LSD contrast tested for a dyadic self-fulfilling effect involving one perceiver and one target by comparing the amount of candy taken by targets in the no- and single-overweight expectation conditions. No significant difference emerged, indicating that individual perceivers' stereotypic expectations did not have a self-fulfilling effect, ($M = 3.04_{no\text{-}overweight}$, $SD_{no\text{-}overweight} = 2.85$ vs. $M_{single\text{-}overweight} = 2.99$, $SD_{single\text{-}overweight} = 3.62$), $t(238) = 0.08$, $p = .94$, 95% CI [-1.19, 1.29], $d = 0.01$, 95% CI [-0.30, 0.32].

Accumulation. Despite no evidence of a dyadic self-fulfilling prophecy effect, there was still the possibility of an accumulation effect. In fact, accumulation may sometimes be necessary for a stereotype to have any self-fulfilling effect at all. In the current data, for example, it was possible that the signal communicated by the situational affordances was too weak to elicit a self-fulfilling effect when only the expectations of individual perceivers were considered, in which case targets might have ignored or discounted it. To address this, a LSD contrast compared the amount of candy taken by targets in the single- and double-overweight expectation conditions. It showed that targets took significantly more candy in the double-overweight ($M = 4.77$, $SD = 5.07$) than single-overweight ($M = 2.99$, $SD = 3.62$) expectation condition, $t(238) = 2.83$, $p = .005$, 95% CI [0.54, 3.02], $d = 0.45$, 95% CI [0.14, 0.76]. This result supports an accumulation effect because it shows that targets more strongly confirmed the overweight stereotype when two perceivers believed the target was heavy than when only one did.

Pattern of accumulation. Finally, the analyses explored the pattern of the accumulation

effect by testing for concurrent and synergistic accumulation. Concurrent accumulation occurs when multiple perceivers each have a unique, additive self-fulfilling effect on a target's behavior (Jussim et al., 1996). The data would support concurrent accumulation if the degree to which targets confirmed the overweight stereotype increased across the no-, single-, and double-overweight expectation conditions, respectively. As reported above, however, there was no significant difference in the amount of candy taken by targets in the no- and single-overweight expectation conditions, thereby indicating no support for concurrent accumulation.

Synergistic accumulation occurs when the self-fulfilling effects of multiple perceivers are stronger in combination than their additive effects would suggest (Madon et al., 2004). The data would support synergistic accumulation if there was a larger increase in targets' confirmatory behavior between the single- and double-overweight expectation conditions than between the no- and single-overweight expectation conditions. The results reported above confirmed this pattern. The difference between the amount of candy taken in the double- versus the single-overweight expectation conditions was larger in magnitude (d = 0.45, 95% CI [0.14, 0.76]) than the difference between the single- versus the no-overweight expectation conditions (d = 0.01, 95% CI [-0.30, 0.32]), with neither effect size included within the confidence interval of the other.

These results were verified with a step-wise regression analysis using forward selection with two predictors: a linear term that corresponded to the number of perceivers in a triad who believed the target was heavy (i.e., 0, 1, and 2), and a quadratic term that was the square of the linear term (i.e., 0, 1, and 4). A significant and positive linear effect would support the presence of concurrent accumulation. A significant and positive quadratic effect would support the presence of synergistic accumulation. Because forward selection enters variables into the model one at a time, leading with the one that explains the greatest variance in the dependent variable, it

permitted the quadratic term to be considered prior to the linear term, which is an appropriate analytic strategy under some conditions (Rawlings, Pantula, & Dickey, 2001). In the current research it was appropriate because neither accumulation process had theoretical priority and each process could occur in combination with, or in the absence of, the other. In addition, because of multicollinearity, artificially forcing the linear term as the first predictor in the model would bias the test against the quadratic term. The results indicated that the quadratic term explained a significant proportion of variance in the amount of candy taken, $t = 3.15$, $p = .002$, whereas the linear term did not meet the entry criterion, $t = .85$, $p = .398$. This result confirms the presence of synergistic, but not concurrent, accumulation.

Discussion

The results of Experiment 1 supported the hypothesis that the self-fulfilling effect of stereotypes can accumulate across perceivers. In Phase 1, individual perceivers selected bins of candy that afforded the target a greater opportunity to confirm the overweight stereotype when they believed the target was heavy versus thin. In Phase 2, targets received a single bowl of candy that combined the candy selections of two independent perceivers whose beliefs about the target's weight had been systematically varied. Because of the way individual perceivers treated the target in Phase 1, this procedure created a situation in which targets' opportunity to confirm the stereotype in Phase 2 was lowest in the no-overweight expectation condition, intermediate in the single-overweight expectation condition, and greatest in the double-overweight expectation condition. However, targets did not more strongly confirm the stereotype in the single- than no-overweight expectation condition despite a greater affordance to do so, thus providing no evidence of the stereotype's self-fulfilling effect within dyadic relations.

Nevertheless, there was evidence of the stereotype's cumulative self-fulfilling effect.

Targets behaviorally confirmed the overweight stereotype more strongly in the double- than single-overweight expectation condition. One might wonder, however, whether this result truly reflects the accumulation of the stereotype's self-fulfilling effect or a mere artifact of the procedures. That is, didn't the tendency for perceivers to disproportionately select bins that contained two (Bin 1) and four (Bin 2) pieces of candy for a target whom they believed was thin create a ceiling effect that limited how much candy targets in the no- and single-overweight expectation conditions could take? And, if this happened, doesn't it mean that these targets had no choice but to confirm the overweight stereotype less than targets in the double-overweight expectation condition? The answer to both questions is *no*.

The greatest potential for a ceiling effect was in the no-overweight expectation condition because it was here that targets received the least amount of candy overall. However, even in this condition targets received more candy than they elected to take. On average, these targets took about three pieces of candy even though nearly half of them received 22 pieces of candy or more, and none received less than 6 pieces. Hence, the amount of candy targets in the no-overweight expectation condition received did not severely restrict how much candy they could take. In fact, had they wanted to, they all could have taken the average amount of candy taken by targets in the double-overweight expectation condition (4.99 pieces), though generally they did not do so, and nearly half of them could have taken the maximum amount of candy taken by any target in the double-overweight expectation (20 pieces), though none did.

In addition, if a ceiling effect had been responsible for the results, then targets in the no-overweight expectation condition should have taken less candy than targets in the single-overweight expectation condition. But, that did not happen either. Despite that targets in the no-overweight expectation condition received an average of 10 fewer pieces of candy than targets in

the single-overweight expectation condition, the amount of candy targets took in these conditions was nearly identical. Finally, in no case were targets compelled to confirm the stereotype. Even in the double-overweight expectation condition targets could have chosen not to take any candy. The fact that few targets made this choice indicates that targets in this condition confirmed the stereotype even though they did not have to. On empirical grounds, therefore, we conclude that ceiling effects cannot explain away the support we obtained for accumulation.

Theoretical considerations also lead us to conclude that the results reflected a bona fide accumulation effect. The amount of candy targets received across the expectation conditions mattered. Perceivers, acting on their stereotypic expectations, constructed a situation that afforded some targets greater opportunity to confirm the stereotype than others. Although this means that targets' confirmatory behavior was somewhat dependent on perceivers' behavior, this is no artifact; this is precisely the process hypothesized to create a self-fulfilling prophecy. According to all theoretical accounts, perceivers treating targets in line with their false expectations is a necessary step of the self-fulfilling prophecy process (e.g., Darley & Fazio, 1980). Thus, the dependency that existed between perceivers' and targets' behavior in this experiment is an inherent component of the underlying self-fulfilling prophecy process and does not invalidate the support we found for accumulation.

It is possible, however, that the dependency contributed to the pattern of synergistic accumulation that characterized the data. The amount of candy targets received increased across the no-, single-, and double-overweight expectation conditions, but differences between the conditions were not equal. On average, 10 pieces of candy separated the no- and single-overweight expectation conditions, whereas 15 pieces separated the single- and double-overweight expectation conditions. Because of the inherent dependency between perceivers' and

targets' behavior, it is possible that the upward trend in perceivers' behavior caused the synergistic pattern of accumulation that was present in targets' confirmatory behavior. Stated differently, it is conceivable that targets' confirmatory behavior might have evidenced concurrent accumulation had perceivers' behavior been linear.

The implication of this interpretation is that targets were passive recipients of the situational affordances with which they were provided. Although the present data cannot rule out this possibility, targets are typically conceptualized as active players in the self-fulfilling prophecy process (Snyder & Stukas, 1999). In the current research, targets may have actively contributed to the stereotype's cumulative self-fulfilling effect through their construal of the situational affordance. For example, targets in the double-overweight expectation condition may have presumed that they could take many pieces of candy without detection, or reasoned that there was no need to refrain since there was plenty of candy to go around. Conversely, targets in the no- and single-overweight expectation conditions might have assumed that any candy they took would likely be noticed, or felt that they should only take only a few pieces so that others could have some too. Though speculative, the idea that targets contributed to the stereotype's synergistic effects via construal processes fits current conceptualizations of the self-fulfilling prophecy, as well as a large body of psychological research showing that people actively interpret social reality (Darley & Fazio, 1980; Fiske 2004; Snyder & Stukas, 1999).

Finally, even though we observed synergistic accumulation with a laboratory procedure that involved count-based behaviors, we are disinclined to attribute the effect solely to the method. The reason for our skepticism is that Madon et al. (2004) observed similar results in a naturalistic study involving parents and their adolescent children. We are not arguing that the same underlying process necessarily operated in both studies. Because the two studies used

different methods, several processes that were prevented from operating in the current research could have operated in Madon et al.'s study. Nevertheless, the fact that both studies found evidence of synergistic accumulation does suggest that the effect is not merely a laboratory phenomenon dependent on count-based behaviors.

Overview of Experiments 2 and 3

Experiments 2 and 3 tested the accumulation hypothesis with respect to sex stereotypes, and did so with a method first used in the self-fulfilling prophecy research of Word et al. (1974). To summarize, Word et al. performed two experiments that tested the self-fulfilling effect of racial stereotypes on the interview performance of job applicants. In the first experiment, White participants interviewed either an African American or a White confederate who posed as a job applicant. Consistent with the idea that stereotypes can bias perceivers' behavior toward targets, participants used a less favorable interview style when interviewing the African American than White confederate.

In the second experiment, White participants played the role of the job applicant and White confederates played the role of the interviewer. The interviewers were trained to use either the less favorable interview style that participants in the first experiment used when interviewing the African American confederate, or the more favorable interview style that participants in the first experiment used when interviewing the White confederate. The results showed that participants in the second experiment performed worse during the interview when they were subjected to the less favorable than more favorable interview style. Thus, the way that the African American confederates had been treated in the first experiment undermined the interview performance of participants in the second experiment.

The method Word et al. (1974) used is referred to as a *double-randomization design*

because it involved the sequential manipulation of two variables (MacKinnon, Fairchild, & Fritz, 2007). First, Word et al. manipulated a job applicant's race to test whether racial stereotypes influenced perceivers' interview style. Second, they manipulated perceivers' interview style to test whether it influenced the interview performance of job applicants. Double-randomization designs are valued because they experimentally test the effect of a mediator variable, thereby ruling out alternative explanations for the link between a prior manipulation and a subsequently measured dependent variable (MacKinnon et al., 2007; MacKinnon, Cheong, & Pirlott, 2012). With this strength in mind, we used a double-randomization design in Experiments 2 and 3. In Experiment 2, we manipulated whether perceivers developed stereotype-consistent expectations about target to test whether it influenced their tendency to provide the target with a situational affordance that encouraged sex-typed behavior. In Experiment 3, we manipulated the number of perceivers who provided a target with a situational affordance that encouraged sex-typed behavior to test whether sex stereotypes had cumulative self-fulfilling effects.

Experiment 2

Experiment 2 constituted the first half of our double-randomization design. It tested whether sex stereotypes can lead perceivers to provide targets with situational affordances that encourage confirmation of the stereotype. Participants read about a fictitious target who was described as a sex-typed woman, a gender-neutral individual, or a sex-typed man. Afterwards, participants selected articles for the target to read from an article pool that included a mix of sex-typed and gender-neutral articles. Participants' article selections indexed the degree to which they provided the target with a situational affordance that encouraged sex-typed behavior.

Participants

Undergraduates (N = 123) at Rutgers University participated to fulfill a course

requirement. There were 52 women and 71 men, including 10 African Americans, 29 Asians/Asian Americans, 64 European Americans, 9 Latina/os, and 11 participants who self-described as “other”.

Experimental Design

Participants were randomly assigned to one of three expectation conditions. In the *feminine expectation condition* (*n* = 42), participants read a profile about a stereotypical woman. In the *gender-neutral expectation condition* (*n* = 43), participants read a profile about a gender-neutral, sex-unidentified person. In the *masculine expectation conditio*n (*n* = 38), participants read a profile about a stereotypical man. Although participants believed the profile described another participant who would be their partner during the experiment, the partner was actually fictitious. Accordingly, participants acted as perceivers, and the fictitious partner was the target.

Materials and Measures

Profile. A handwritten profile described the target's name, age, hobbies and employment. In the feminine expectation condition, the profile described a sex-typed target named Jessica who enjoyed gymnastics and going shopping with friends, and who worked as a baby-sitter in high school and at a cosmetics counter in a department store during the summer. In the gender-neutral expectation condition, the profile described a gender-neutral target named Jesse who enjoyed swimming and hanging out with friends, and who worked as a cashier at a local grocery store in high school and as a lifeguard during the summer. In the masculine expectation condition, the profile described a sex-typed target named Michael who enjoyed playing lacrosse and video games, and who worked as a landscaper in high school and at a gas station during the summer.

Perceiver behavior. To assess whether perceivers provided the target with a situational affordance that encouraged confirmation of sex stereotypes, perceivers selected three magazine

and newspaper articles for the target to read. Perceivers selected these articles from one of two article pools. We used two article pools rather than one to prepare for the procedures of Experiment 3 which, as we describe later, required two, separate and different sets of articles in order to simulate the behavior of two independent perceivers. The article pools both presented the titles and brief summaries of nine articles, including three stereotypically feminine articles, three gender-neutral articles, and three stereotypically masculine articles that were matched for length. The stereotypically feminine articles dealt with fashion, relationships, and health and beauty. The gender-neutral articles dealt with travel, food, and entertainment. The stereotypically masculine articles dealt with mechanics, science, and space exploration. We assigned a value of +1 to each selected stereotypically feminine article, a value of 0 to each selected gender-neutral article, and a value of -1 to each selected stereotypically masculine article, and then summed these coded values to create a single value per perceiver. Higher values indicated that perceivers selected a greater number of stereotypical feminine articles for the target to read. We subsequently refer to this variable as *article selection*.

Manipulation check. Two trait judgments assessed whether the expectation manipulation induced stereotype-consistent expectations: "*How masculine vs. feminine is your partner?"* and *"How manly vs. womanly is your partner?"* Perceivers responded on 7-point scales with anchors 1 (*very masculine; very manly*), 4 (*gender-neutral*) and 7 (*very feminine; very womanly)*. To reduce suspicion, these questions were embedded among fillers that assessed non-sex-typed trait judgments and interests. Responses to the two sex-typed judgments were averaged to create a new variable (α = .68). Higher values indicated greater perceived femininity.

Suspicion check. Perceivers reported what they believed the experiment was about.

Procedures

After obtaining informed consent, the experimenter described the study as an exploration into how well people can predict the likes, dislikes, and abilities of another person about whom they know little. Perceivers expected to meet with another participant who would be their partner during the study. In anticipation of the meeting, perceivers completed a survey that assessed their name, age, hobbies and employment under the guise that their responses would be given to their partner (i.e., the fictitious target) whom they believed was completing the same survey in another room. After completing the survey, the experimenter collected it, exited the room, and returned with what appeared to be their partner's survey, but which was actually the target's profile. Perceivers were told to study their partner's survey because they would use it to select activities for their partner to perform. Perceivers then received an article pool from which they selected three articles that they believed their partner would most enjoy reading. The experimenter explained that the perceiver and partner would later discuss how much the partner actually enjoyed reading the selected articles. Perceivers subsequently completed the manipulation and suspicion checks, reported demographic information, and were debriefed.

Results

Preliminary Analyses

Descriptive statistics. Supplementary Table 3 presents descriptive statistics.

Manipulation check. A one-way ANOVA followed by three Tukey HSD contrasts tested whether the expectation manipulation induced perceivers to hold stereotype-consistent expectations about the target. The independent variable was the expectation (feminine vs. gender-neutral vs. masculine). The dependent variable was perceivers' judgments of the target's sex-typed traits. The ANOVA results and pattern of means supported the manipulation's

effectiveness. Perceivers judged the target as most feminine in the feminine expectation condition (M = 4.88, SD = 1.12), intermediately feminine in the gender-neutral expectation condition (M = 4.25, SD = 0.94), and least feminine in the masculine expectation condition (M = 2.86, SD = 1.01), $F(2, 120) = 42.90$, $p < .001$, $\eta^2 = .42$, 90% CI [.30, .50]. These values all differed significantly from one another, ts(120) $\geq$ 2.74, all ps $\leq$ .019, all ds $\geq$ .61.

Suspicion check. There was one suspicious perceiver. Removing this perceiver's data did not meaningfully alter the results, and the data were retained in all of the analyses.

Article pools. A 2 (article pool: pool one vs. pool two) × 3 (expectation condition: feminine vs. gender-neutral vs. masculine) ANOVA tested whether the two article pools differentially influenced perceivers' article selections. Because neither the main effect of article pool nor the interaction between article pool and expectation was significant, Fs(1, 117) $\leq$ 1.59, ps $\geq$ .209, both $\eta_p^2 \leq 0.03$, we omitted article pool as a factor in the main analyses.

Main Analyses

A one-way ANOVA followed by three Tukey HSD contrasts tested whether perceivers afforded the target a greater opportunity to confirm sex stereotypes when the target was sex-typed versus gender-neutral. The independent variable was the expectation (feminine vs. gender-neutral vs. masculine). The dependent variable was article selection. The results showed that the number of stereotypically feminine articles selected by perceivers was highest in the feminine expectation condition (M = 1.64, SD = 1.48), intermediate in the gender-neutral expectation condition (M = .16, SD = 1.28), and lowest in the masculine expectation condition (M = -2.00, SD = 0.87), $F(2, 120) = 93.70$, $p < .001$, $\eta^2 = .61$, 90% CI [.52, .67], with these values all differing significantly from one another: Feminine vs. gender-neutral expectation conditions, $t(120) = 5.38$, $p < .001$, 95% CI [0.83, 2.14], $d = 1.20$, 95% CI [0.73, 1.67]; Feminine vs.

masculine expectation conditions, $t(120) = 13.61$, $p < .001$, 95% CI [3.01, 4.28], $d = 2.95$, 95% CI [2.38, 3.51]; Gender neutral vs. masculine expectation conditions, $t(120) = 7.85$, $p < .001$, 95% CI [1.51, 2.81], $d = 1.75$, 95% CI [1.26, 2.23]. Table 2 reports the average number of stereotypically feminine, gender-neutral, and stereotypically masculine articles selected in each expectation condition.

Because the number of women and men was not equal across the expectation conditions, we also examined whether perceivers' sex could account for the results, but found no evidence to support its influence. Neither the main effect of perceiver sex nor the *Expectation* × *Perceiver sex* interaction had a significant effect on perceivers' article selections, Fs ≤ 2.36, ps $\geq .127$, all $\eta_p^2 \leq .03$, and in no case did the inclusion of perceiver sex meaningfully alter either the pattern or significance of the expectation's effect. We also analyzed the data separately for women and men and found virtually identical results to those produced by the full sample.

Discussion

Perceivers in this experiment selected the greatest number of stereotypically feminine articles for Jessica, the sex-typed feminine target, the greatest number of stereotypically masculine articles for Michael, the sex-typed masculine target, and an intermediate number of sex-typed articles for Jesse, the gender-neutral target. These results clearly show that perceivers afforded the sex-typed targets greater opportunity to confirm sex stereotypes than they afforded the gender-neutral target. However, the cause of this effect is less clear. Because the procedures paired a sex-typed name with stereotype-consistent attributes in the feminine and masculine expectation conditions, perceivers' article selections could have been influenced by sex stereotypes, or the target's stereotype-consistent attributes. If it is the latter, then perhaps Experiment 2 is more about target-based expectations than stereotype-based expectations. For

three reasons, we do not believe this to be the case.

First, the congruence between the target's sex and the target's stereotype-consistent attributes likely provided perceivers with strong justification to apply their stereotypes when making their article selections (Crandall & Eshleman, 2003; Fazio, 1990; Madon et al., 2006). Second, the amount of stereotype-consistent attributes provided to perceivers was rather modest, which likely limited its influence even if perceivers were motivated to form target-based impressions. Third, we observed the same pattern in Experiment 1 even though the target's attributes were held constant across the expectation conditions. These considerations suggest that perceivers in the current experiment likely relied on sex stereotypes to make their article selections. However, because this experiment cannot tease apart the relative influence of the target's sex and the target's attributes, it remains possible that different results might have emerged had the procedures varied only the target's sex.

Experiment 3

Experiment 3 constituted the second half of our double-randomization design. Whereas Experiment 2 showed that individual perceivers provided sex-typed targets with a situational affordance that encouraged confirmation of sex stereotypes, Experiment 3 examined whether these affordances, if imposed by multiple perceivers, can cause the stereotype to have cumulative self-fulfilling effects. The experiment manipulated the number of perceivers who provided participants with a sex-typed situational affordance. Participants received a set of articles that they were told had been chosen for them by two participants who would be their partners during the experiment. The number of sex-typed and gender-neutral articles in the sets was varied to simulate the behavior of two perceivers from Experiment 2. The number of sex-typed articles participants selected to read indexed the degree to which they confirmed sex stereotypes.

Participants

Undergraduates (*N* = 121) at Rutgers University participated to fulfill a course requirement. There were 65 women and 56 men. Race was not assessed.

Design

Participants were randomly assigned to one of five conditions that manipulated the number of perceivers who provided them with a sex-typed situational affordance. In two conditions, two perceivers provided participants with a situational affordance that encouraged confirmation of the stereotype of women (*double-feminine*, *n* = 26) or men (*double-masculine*, *n*= 25). In two other conditions, one perceiver provided participants with a situational affordance that encouraged confirmation of the stereotype of women (*single-feminine*, *n*= 25) or men (*single-masculine*, *n*= 20), whereas the other perceiver provided a situational affordance that discouraged confirmation of sex stereotypes. In the *gender-neutral* condition (*n* = 25), two perceivers provided participants with a situational affordance that discouraged confirmation of sex stereotypes. To manipulate these affordances, participants received one of five article sets that included different numbers of sex-typed and gender-neutral articles, and selected three to read. Hence, the participants acted as targets, and even though the two perceivers with whom they were paired were fictitious, the number of sex-typed and gender-neutral articles included in the article sets were based on the average article selections of perceivers in Experiment 2.

Materials and Measures

Article sets. Five article sets manipulated the situational affordances perceivers provided targets. Each article set showed the titles and brief summaries of six magazine and newspaper articles. The six articles in each set simulated the behavior of two perceivers, each of whom ostensibly had selected three articles for the target to read. We varied the number of sex-typed

articles in the sets based on the results of Experiment 2 (Table 2). Specifically, we rounded the average article selections of perceivers in the feminine and masculine expectation conditions of Experiment 2 to the nearest integer, and then paired these rounded values in different permutations to simulate the behavior of two perceivers combined. We set the remaining articles to be gender-neutral.

For example, consider the article sets used in the double- and single-feminine conditions. Table 2 reports that perceivers in the stereotypically feminine expectation condition of Experiment 2 selected an average of two stereotypically feminine articles and 0.36 stereotypically masculine articles. To create the article set for the double-feminine condition in Experiment 3, we rounded these values to two and zero, respectively, and then doubled them to create an article set that included four stereotypically feminine articles and zero stereotypically masculine articles. We added two gender-neutral articles to complete the set of six. This article set simulated the behavior of two independent perceivers, both of whom held a stereotypic expectation about the target.

We did not double the rounded values when creating the article set for the single-feminine condition because it was intended to simulate the behavior of a single perceiver who held a stereotypic expectation about the target. Thus, it included two stereotypically feminine articles, zero stereotypically masculine articles, plus four gender-neutral articles to complete the set of six. The same procedures created the remaining article sets: (1) double-masculine: zero stereotypically feminine articles, four stereotypically masculine articles, and two gender-neutral articles; (2) single-masculine: zero stereotypically feminine articles, two stereotypically masculine articles, and four gender-neutral articles, and; (3) gender-neutral: one stereotypically feminine article, one stereotypically masculine article, and four gender-neutral articles.

We used these procedures for three reasons. First, because perceivers in Experiment 2 could not select a fraction of an article for the target to read, rounding the average values to the nearest integer best represented the behavior a single perceiver acting independently. Likewise, combining the rounded values best represented the combined behavior of two perceivers acting independently. Second, because the rounded values always equaled the modal response, they also best represented the typical behavior of individual perceivers in the stereotypically feminine and stereotypically masculine expectation conditions of Experiment 2. Finally, we set all of the remaining articles in the article sets to be gender-neutral so that all targets in the current experiment, even those in the double-feminine and double-masculine conditions, could predominantly disconfirm sex stereotypes.

In considering these procedures, it might seem that targets in the current experiment should have had stereotypically masculine articles to select in the double- and single-feminine conditions, and stereotypically feminine articles to select in the double- and single-masculine conditions. In fact, however, the article sets represented a faithful transmission of the restricted situational affordances perceivers in Experiment 2 had provided the sex-typed targets. Although we could have used less restricted situational affordances than did perceivers in Experiment 2, doing so would have disrupted a key step of the self-fulfilling prophecy process – i.e., perceivers treating targets in line with their false expectations – and undermined the effect that we were trying to capture in the second half of our double-randomization design. The procedure we used to create the article sets avoided this problem.

Target behavior. Targets selected three articles to read from the six article set.

Profile. Targets completed a profile about themselves (name, age, hobbies, employment) that they believed would be given to their partners, reinforcing the cover story that the six

articles in the set had been given to them by their partners.

Suspicion check. Targets described what they believed the experiment was about.

Procedures

The procedures matched those from Experiment 2 with these exceptions. Targets expected to be paired with two other participants who would be their partners during the experiment. Targets were told that their partners were using their profiles to design activities for them, the first of which involved reading three magazine and newspaper articles from a set of six articles that had been chosen for them by their partners from a larger pool of articles on a variety of topics. Targets expected to discuss how much they enjoyed reading the articles with their partners later in the experiment. Targets then received the six articles that their partners had supposedly chosen for them, and selected three to read. Afterward, targets completed the suspicion check, provided demographic information, and were then debriefed.

Results

Preliminary Analyses

Descriptive statistics. Supplementary Table 4 presents descriptive statistics.

Suspicion check. There were four suspicious targets. Removing their data from the analyses did not meaningfully alter the results, and their data were retained in all of the analyses.

Main Analyses

Perceivers in Experiment 2 provided sex-typed targets with a situational affordance that encouraged confirmation of sex stereotypes. In the current research, a series of analyses tested whether these situational affordances, if imposed by multiple perceivers, can cause sex stereotypes to have cumulative self-fulfilling effects on targets' behavior. To allow for the possibility that the stereotypes of women and men might evidence different effects, the analyses

separately analyzed the feminine and masculine conditions.

Participant behavior. Two, separate, one-way ANOVAs tested whether the number of perceivers who provided targets with a sex-typed situational affordance influenced how strongly targets confirmed sex stereotypes. One ANOVA focused on the stereotype of women. It examined how many stereotypically feminine articles targets selected in the double-feminine, single-feminine, and gender-neutral conditions. The other ANOVA focused on the stereotype of men. It examined how many stereotypically masculine articles targets selected in the double-masculine, single-masculine, and gender-neutral conditions. Both ANOVAs were significant: for the stereotype of women, $F(2, 73) = 23.28$, $p < .001$, $\eta^2 = .39$, 90% CI [.23, .49]; for the stereotype of men, $F(2, 67) = 50.11$, $p < .001$, $\eta^2 = .60$, 90% CI [.46, .68]. These results indicate that the number of perceivers who provided targets with a sex-typed situational affordance influenced how strongly targets confirmed sex stereotypes.

Self-fulfilling prophecy effect. Next, two LSD contrasts tested whether the stereotypes of women and men had dyadic self-fulfilling effects. The contrast relevant to the stereotype of women compared the number of stereotypically feminine articles selected by targets in the gender-neutral and single-feminine conditions. The contrast relevant to the stereotype of men compared the number of stereotypically masculine articles selected by targets in the gender-neutral and single-masculine conditions. The results indicated that targets selected significantly more stereotypically feminine articles in the single-feminine than gender-neutral condition, $t(73) = 4.41$, $p = .001$, 95% CI [0.44, 1.16], $d = 1.25$, 95% CI [0.65, 1.83], and significantly more stereotypically masculine articles in the single-masculine than gender-neutral condition, $t(69) = 2.34$ $p = .022$, 95% CI [0.06, 0.80], $d = 0.70$, 95% CI [0.10, 1.30]. Table 3 presents the means. Thus, targets confirmed sex stereotypes more strongly when one perceiver provided them with a

sex-typed situational affordance than when no perceiver had done so. These results replicate prior research showing that stereotypes can have self-fulfilling effects within dyadic relations (Snyder et al., 1977; Word et al., 1974).

Accumulation. Two additional LSD contrasts tested whether the stereotypes of women and men had cumulative self-fulfilling effects. One contrast compared the number of stereotypically feminine articles selected by targets in the single- and double-feminine conditions. The other contrast compared the number of stereotypically masculine articles selected by targets in the single- and double-masculine conditions. The results showed that targets selected significantly more stereotypically feminine articles in the double-feminine than single-feminine condition, $t(73) = 2.27$, $p = .026$, 95% CI [0.05, 0.77], $d = 0.64$, 95% CI [0.08, 1.19], and significantly more stereotypically masculine articles in the double-masculine than single-masculine condition, $t(67) = 6.80$, $p < .001$, 95% CI [0.88, 1.62], $d = 2.04$, 95% CI [1.35, 2.72]. Table 3 presents the means. These results support the accumulation hypothesis because they show that targets confirmed sex stereotypes more strongly when two perceivers provided them with a sex-typed situational affordance than when only one did.

Pattern of accumulation. Additional analyses explored the pattern of these accumulation effects. For the stereotype of women, the data would support concurrent accumulation if targets selected an increasingly greater number of stereotypical feminine articles across the gender-neutral, single-feminine, and double-feminine conditions, respectively. Parallel increases in targets' selection of stereotypically masculine articles across the gender-neutral, single-masculine, and double-masculine conditions would support concurrent accumulation for the stereotype of men. The fact that the contrasts reported above supported the significance of these comparisons suggests that concurrent accumulation was present for both stereotypes.

Because concurrent and synergistic accumulation can occur simultaneously, the analyses also tested whether the stereotypes evidenced synergistic accumulation. The results would support synergistic accumulation for the stereotype of women if the difference between the number of stereotypically feminine articles selected by targets in the double-feminine versus single-feminine conditions was larger in magnitude than the difference between the number of stereotypically feminine articles selected by targets in the single-feminine versus gender-neutral conditions. A parallel pattern of differences in targets' selection of masculine articles in the gender-neutral, single-masculine, and double-masculine conditions would support the presence of synergistic accumulation for the stereotype of men.

The results did not support synergistic accumulation for the stereotype of women. In fact, the pattern was in the opposite direction. As reported above, the magnitude of the difference between the double- and single-feminine conditions was smaller (d = 0.64, 95% CI [0.08, 1.19]) than the magnitude of the difference between the single-feminine and gender-neutral conditions (d = 1.25, 95% CI [0.65, 1.83]). A forward selection regression analysis confirmed this result. Whereas a linear term corresponding to the number of perceivers who provided targets with a feminine sex-typed situational affordance (i.e., 0, 1, and 2) explained a significant proportion of variance in the number of stereotypically feminine articles selected by targets, t = 6.68, p < .001, a quadratic term that was the square of the linear term (i.e., 0, 1, and 4) did not meet the entry criterion, t = 1.25, p = .214. Thus, the stereotype of women evidenced concurrent, but not synergistic, accumulation.

By contrast, the stereotype of men did evidence synergistic accumulation. As reported above, the magnitude of the difference between the double- and single-masculine conditions was larger (d = 2.04, 95% CI [1.35, 2.72]) than the magnitude of the difference between the single-

masculine and gender-neutral conditions (d = 0.70, , 95% CI [0.10, 1.30]), with neither effect size included within the confidence interval of the other. Moreover, a forward-selection, step-wise regression analysis that included a linear term corresponding to the number of perceivers who provided targets with a masculine sex-typed situational affordance (i.e., 0, 1, and 2), and a quadratic term that was the square of the linear term (i.e., 0, 1, and 4) indicated that the quadratic term explained a significant proportion of variance in the number of stereotypically masculine articles selected, t = 10.09, p < .001, whereas the linear term did not meet the entry criterion, t = .60, p = .953. This result helps to clarify the contrast results reported above for the stereotype of men; although differences between the means suggested the presence of concurrent accumulation, those differences were predominantly driven by a strong quadratic effect indicative of synergistic accumulation. Synergistic accumulation, therefore, best characterized the stereotype of men.

Participant sex. Because there were not equal numbers of women and men in the conditions, target sex was added to the ANOVAs described above to test whether it could account for the results. The main effect of target sex was significant, with women selecting significantly more stereotypically feminine articles than men (Ms = 1.43 vs. 1.11) and fewer stereotypically masculine articles than men (Ms = 0.88 vs. 1.73), Fs $\geq$ 14.05, ps < .001, both $\eta_p^2 \geq .17$. However, target sex never interacted with the experimental manipulation, Fs $\leq$ 1.37, ps $\geq$ .260, both $\eta_p^2 \leq .04$, nor did its inclusion meaningfully alter the manipulation's effect. We also reanalyzed the data separately for women and men and found virtually identical results to those produced by the full sample.

Discussion

Experiment 3 showed that targets confirmed sex-stereotypes more strongly when two

perceivers provided them with a sex-typed situational affordance than when only one had. Because the results were the product of a double-randomization design, they also showed that a main reason targets in Experiment 3 confirmed sex stereotypes was because they were encouraged to do so by the situational affordances that perceivers in Experiment 2 constructed for them. Indeed, a primary way that perceivers are hypothesized to elicit confirmatory behavior from targets is by constructing situations that afford them opportunities and constraints that encourage an expectancy-consistent response. Our results were consistent with this hypothesized mechanism.

Of course, if the situational affordances were so restrictive that they prevented targets from disconfirming the stereotypes, then the observed effects would not be particularly interesting. However, this was not the case. All targets in Experiment 3 could have predominantly disconfirmed sex stereotypes. For example, targets in the single-feminine and single-masculine conditions could have chosen all gender-neutral articles, thus avoiding the sex-typed articles entirely, and targets in the double-feminine and double-masculine conditions were not required to select more than a single sex-typed article. In no case, therefore, were targets compelled to confirm sex stereotypes. Nevertheless, they did confirm sex stereotypes and, critically, they did so *more than* the situational affordances required.

Although the data cannot explain why the situational affordances had this effect, it may have to do with the way that targets construed the situation (Darley & Fazio, 1980; Snyder & Stukas, 1999). Targets believed that their partners had chosen articles for them on the basis of their profiles, and expected to discuss with their partners how much they enjoyed reading a subset of them. Thus, in targets' minds, the articles appeared to reflect their partners' perceptions of them, and the articles they chose to read appeared to have interpersonal consequences. Given

these parameters, targets might have gone along with the situational affordances for two reasons.

First, targets' belief that their partners selected articles for them to read on the basis of their profiles may have led them to engage in a process akin to the confirmation bias; they may have read the article summaries with an eye toward seeing how their interests aligned with the articles, a strategy that may have caused them to see connections that they otherwise would have missed. In turn, these connections may have increased their willingness to read sex-typed articles that they might have rejected under different circumstances. Second, targets may have wanted to affiliate with their partners in the anticipated meeting and perceived this goal to be more achievable if they chose the most prevalent type of article in the article sets. For example, targets in the double-feminine condition may have selected to read more stereotypically feminine articles than required to make the anticipated meeting go more smoothly. This interpretation is consistent with the established finding that targets are more susceptible to self-fulfilling prophecies when they are motivated to get along with perceivers (Snyder & Stukas, 1999).

Experiment 4

Experiments 1, 2, and 3 supported the hypothesis that the self-fulfilling effect of stereotypes can accumulate across perceivers. Perceivers treated targets in stereotypic ways, and the greater the number of perceivers who did so, the more strongly targets' confirmed a stereotype. We performed Experiment 4 to examine a potential downstream consequence of this effect: Once the cumulative self-fulfilling effect of a stereotype has been set in motion, it may influence the impressions that new perceivers form about a target, thereby bringing the self-fulfilling prophecy process full circle. One process through which this could occur is the fundamental attribution error (Jones & Davis, 1965; Ross, 1977).

Perceivers do not always take situational factors into account as much as they should

when trying to understand the causes of a target's behavior. Instead, they tend to assume that a target's behavior corresponds to an underlying disposition. Perceivers' tendency to discount situational factors in favor of correspondent inferences may perpetuate a stereotype's cumulative self-fulfilling effect in the following way: New perceivers who observe a target behave in a stereotypic way may misattribute the behavior to an underlying disposition when it was actually caused by a stereotype's cumulative self-fulfilling effect. Such a tendency might even occur among perceivers who are aware that prior perceivers encouraged the target's stereotypic behavior. After all, some of the earliest attribution research established that perceivers make correspondent inferences about targets who they know lacked free choice (Jones & Harris, 1967). This finding raises the possibility that new perceivers will assume that targets' behavior reflects their dispositions even when they know that the behavior was shaped by the actions of others. Consistent with this possibility, we tested two interrelated hypotheses in Experiment 4.

One hypothesis was that a stereotype's self-fulfilling effect can influence the way new perceivers judge targets. Although this is not the first time such an effect has been examined (Snyder & Swann, 1978), it does underscore a theoretically important point that has been overlooked; perceivers who judge targets solely on the basis of individuating information (a process referred to as *individuation*) may develop stereotypic impressions when targets' individuating information was caused by a stereotype's self-fulfilling effect (Jussim, 2012). Thus, individuation (which, by definition, involves no stereotyping) can cause perceivers to form stereotypic impressions, thereby setting the stage for future restrictions on targets' opportunities.

The other hypothesis was that perceivers do not appropriately adjust their impressions even when they know that targets' stereotypic behavior was shaped by the actions of others. This hypothesis represents a novel test of the fundamental attribution error because it conceptualizes it

as a mechanism through which a stereotype's cumulative self-fulfilling effect can transcend social interactions. We tested these hypotheses with respect to sex stereotypes, focusing on the common beliefs that women and men have different personality traits, academic competencies, career aptitudes, and suitability for careers (Haines, Deaux, & Lofaro, 2016).

Method

Participants

Undergraduates (*N* = 230) at Rutgers University participated to fulfill a course requirement. There were 138 women, 90 men, and 2 participants who omitted a response, including 16 African Americans, 65 Asians/Asian Americans, 112 European Americans, 15 Latina/os, 19 participants who reported "other", and 3 participants who omitted a response.

Design

Participants were randomly assigned to a 5 (target behavior: strong-feminine vs. moderate-feminine vs. gender-neutral vs. moderate-masculine vs. strong-masculine) × 2 (situational awareness: aware vs. unaware) between-subjects experimental design. Target behavior manipulated how strongly a target confirmed sex stereotypes by varying the number of stereotypically feminine and masculine articles the target selected to read from a group of articles that afforded sex-typed behavior to varying degrees. Situational awareness manipulated whether participants were made aware of the situational affordances that gave rise to the target's article selections. Although the target was fictitious, participants believed the target would be their partner during the experiment, and the target's article selections, though experimentally manipulated, were based on the average article selections of participants in Experiment 3. Therefore, participants acted as perceivers and the fictitious partner was the target.

Materials and Measures

Article selection lists. Five article selection lists manipulated target behavior. Each list included the titles and brief summaries of three magazine and newspaper articles that the target had purportedly selected to read. The number of stereotypically feminine, gender-neutral, and stereotypically masculine articles in the lists matched the average number of stereotypically feminine, gender neutral, and stereotypically masculine articles, rounded to the nearest integer, that had been selected by targets in the five conditions of Experiment 3. For example, Table 3 shows that targets in the double-feminine condition of Experiment 3 chose an average of 1.81 stereotypically feminine articles, 1.15 gender-neutral articles, and zero stereotypically masculine articles. These values were rounded to the nearest integer to create the article selection list for the strong-feminine condition in the current experiment: two stereotypically feminine articles, one gender-neutral article, and zero stereotypically masculine articles.

The same procedure created the remaining article selection lists: (1) strong-masculine condition: zero stereotypically feminine articles, one gender neutral article, and two stereotypically masculine articles; (2) moderate-feminine condition: one stereotypically feminine article, two gender-neutral articles, and zero stereotypically masculine articles; (3) moderate-masculine condition: zero stereotypically feminine articles, two gender-neutral articles, and one stereotypically masculine article; (4) gender-neutral condition: one stereotypically feminine article, two gender-neutral articles, and one stereotypically masculine article. Note that the absence of counterstereotypic articles in the strong and moderate conditions was a direct consequence of perceivers' behavior in Experiment 2. That is, because perceivers in Experiment 2 overwhelmingly chose not to select any counterstereotypic articles for the sex-typed targets to read, targets in Experiment 3 had none to select in any but the gender-neutral condition. The

behavior of perceivers in Experiment 2, therefore, had a cascading effect that was represented here in the target's article selections.

Articles. The experiment manipulated situational awareness by giving perceivers in the aware and unaware conditions different information about the group of articles from which the target selected three to read. Perceivers in the aware conditions received one of the five article sets from Experiment 3, all of which conveyed the target's restricted choices: In the strong-feminine condition they received the article set from the double-feminine condition of Experiment 3 (four stereotypically feminine articles, two gender-neutral articles, and zero stereotypically masculine articles). In the strong-masculine condition they received the article set from the double-masculine condition of Experiment 3 (zero stereotypically feminine articles, two gender-neutral articles, and four stereotypically masculine articles). In the moderate-feminine and moderate-masculine conditions they received the article sets from the single-feminine (two stereotypically feminine articles, four gender-neutral articles, and zero stereotypically masculine) or single-masculine (zero stereotypically feminine articles, four gender-neutral articles, and two stereotypically masculine articles) conditions of Experiment 3, respectively. In the gender-neutral condition they received the article set from the gender-neutral condition of Experiment 3 (one stereotypically feminine article, four gender-neutral articles, and one stereotypically masculine article). Perceivers in the unaware conditions received the 18-article pool given to perceivers in Experiment 2 (six stereotypically feminine, six gender-neutral, six stereotypically masculine articles). Because these articles reflected diverse interests, they gave the impression that the target had considerable choice when selecting which three articles to read.

Measures

Perceiver judgments. Perceivers judged the target along four dimensions that are

relevant to sex stereotypes, including sex-typed traits, general academic competency, aptitude in the traditionally masculine fields of math and science, and career suitability.

Sex-typed traits. Ten questions assessed perceivers' judgments of the target's sex-typed traits. Two questions used a 7-point bipolar scale to assess how *gentle vs. rough and emotional vs. logical* they judged the target with anchors 1 (*very gentle* or *very emotional*) and 7 (*very rough* or *very logical*). The rest used a 7-point response scale with anchors 1 (*not at all*) and 7 (*very*) to assess the degree to perceivers judged the target to possess four stereotypically feminine traits (*attuned to others feelings, feminine, sensitive, and submissive*), and four stereotypically masculine traits (*confident, masculine, independent, and assertive*). Responses were reverse scored as necessary and then averaged to create one score per perceiver (α = .76). Higher scores reflected the perception that the target possessed more stereotypically feminine traits.

General academic competency. Four questions assessed perceivers' judgments of the target's general academic competency: (1) *How strong is your partner's overall academic performance?* (2) *Would your partner be confident in her or his general academic abilities?* (3) *How intelligent is your partner?, and* (4) *How smart do you consider your partner?* Perceivers responded on 7-point scales with anchors 1 (*not at all*) and 7 (*very*). Responses were averaged to create one score per perceiver (α = .88), with higher values reflecting the perception that the target had greater general academic competency.

Aptitude in math and science. Four questions assessed perceivers' judgments of the target's aptitude in the domains of math and science: (1) *Would your partner be confident in her or his math abilities?* (2) *Would your partner be confident in her or his scientific abilities?* (3) *Would your partner be good at math?*, and (4) *Would your partner be good at science?* Perceivers responded on 7-point scales with anchors 1 (*not at all*) and 7 (*very*). Responses were

averaged to create one score per perceiver (α = .90). Higher values reflected the perception that the target had more aptitude for these traditionally masculine domains.

Career suitability. To assess perceivers' judgments of the target's career-suitability, perceivers answered four questions relevant to careers with sex-typed subspecialties. For example, one question asked *"If your partner was a doctor, would she* or *he be better at surgery or dermatology?"*. Perceivers responded to this question on a 7-point scale with anchors 1 (*surgery*) and 7 (*dermatology*). The other careers and subspecialties were high school teacher (*English teacher vs. science teacher*), trades person (*hairstylist vs. mechanic*), and sales person (*cosmetic sales person vs. car sales person*). The subspecialties of dermatology, English teacher, hairstylist, and cosmetic sales person corresponded to feminine sex-typed subspecialties. Responses were averaged to create one score per perceiver (α = .86). Higher values indicated that perceivers judged the target as better-suited for stereotypically feminine careers.

Manipulation checks. Seven questions assessed whether perceivers had adequately attended to the target's article selections. One question asked perceivers to freely recall the titles of the target's article selections. The other six questions used a multiple-choice format to assess perceivers' memory for the article summaries, which described the three articles selected by the target. Each multiple-choice question had only one correct response. Responses to the multiple-choice questions were coded as 0 (incorrect) or 1 (correct) and then summed to create one variable per perceiver that equaled the total number of correct responses.

Suspicion check. Perceivers reported what they believed the experiment was about.

Procedures

The procedures matched those from Experiment 3 with these exceptions. Perceivers expected to receive information about activities their partner had performed, and to judge their

partner's interests, characteristics, and abilities. Following this cover story, perceivers were told that their partner had already received the titles and brief summaries of a set of magazine and newspaper articles and, from this set, had chosen three articles to read. Perceivers then received information about the articles that had ostensibly been available for their partner to choose, and an article selection list that showed the three articles their partner ultimately selected to read. Perceivers next judged their partner on the four stereotype-relevant dimensions, after which they completed the manipulation and suspicion checks, and were then debriefed.

Results

Preliminary Analyses

Descriptive statistics. Supplementary Table 5 presents descriptive statistics.

Manipulation checks. There were 211 perceivers (93%) who correctly listed all three article titles, 159 (69%) who correctly answered all six multiple choice questions concerning these articles, and 149 (65%) who scored perfectly on both measures. Although a few perceivers correctly listed either zero (n = 9, 4%) or one (n = 3, 1%) article title, these perceivers still performed reasonably well on the multiple choice questions, answering at least four of the six questions correctly. Thus, perceivers had adequately attended to the target's article selections.

Suspicion check. There were four suspicious perceivers. Removing their data did not meaningfully alter the results, and their data were retained in all of the analyses.

Main Analyses

The data were first analyzed with a 5 (target behavior) × 2 (situational awareness) MANOVA in which the dependent variables were judgments of the target's sex-typed traits, general academic competency, aptitude in math and science, and career suitability. There was a substantial main effect for target behavior, *Wilks'* λ = .374, $F(16, 642.20) = 15.25$, $p < .001$, η_p^2

= .22. Neither the main effect of situational awareness, *Wilks'* λ = .992, $F(4, 210) = 0.41$, $p = .800$, $\eta_p^2 = .01$, nor the *Target behavior* × *Situational awareness* interaction, *Wilks'* λ = .944, $F(16, 642.20) = 0.76$, $p = .733$, $\eta_p^2 = .01$, achieved significance. Four ANOVAs replicated these results (Supplementary Table 6), and yielded significant linear relations between target behavior and each dependent variable, *t*s $\geq$ 7.30, *p*s < .001. Tukey HSD contrasts further supported these linear trends (Supplementary Table 7). Table 4 presents the means for target behavior.

These results reveal two clear patterns. First, perceivers judged the target as having more stereotypically feminine characteristics and abilities the greater the number of stereotypically feminine articles selected by the target. Second, this effect was even present among perceivers who were made aware of the target's constrained choices. These results support the hypotheses that a stereotype's cumulative self-fulfilling effect can have ripple effects that influence the way new perceivers judge a target, and that these ripple effects are not mitigated by perceivers' awareness that a target's behavior was shaped by the actions of others.

Because there were not equal numbers of women and men in the ten conditions, perceiver sex was added to the MANOVA described above to examine its effect. There was a significant main effect of perceiver sex, *Wilks'* λ = .951, $F(4, 199) = 2.54$, $p = .041$, $\eta_p^2 = .05$, and a marginally significant *Perceiver sex* × *Target behavior* × *Situational awareness* interaction, *Wilks'* λ = .887, $F(16, 608.59) = 1.52$, $p = .088$, $\eta_p^2 = .03$. Follow-up ANOVAs yielded a significant main effect of perceiver sex: Women judged the target as having fewer stereotypically feminine sex-typed traits than did men (*M*s = 3.94 vs. 4.18), $F(1, 203) = 8.64$, $p = .004$, $\eta_p^2 = .04$. The ANOVAs also yielded two marginally significant interactions: In the unaware condition, women judged the target as having somewhat more general academic

competency than did men, (*M*s = 5.02 vs. 4.70), $F(1, 204) = 2.76$, $p = .098$, $\eta_p^2 = .01$. In the strong-feminine condition, women judged the target as somewhat better-suited for stereotypically feminine careers than did men (*M*s = 6.01 vs. 5.32), $F(4, 206) = 2.15$, $p = .076$, $\eta_p^2 = .04$. In no case, however, did perceiver sex meaningfully alter the effects of target behavior or situational awareness. Additional ANOVAs that examined the data of women and men separately produced results that were nearly identical to those produced by the full sample. Supplementary Figures 1a – 1d present scatter plots of women's and men's judgments across the five experimental conditions of target behavior.

Discussion

The results of Experiment 4 broadly confirmed the hypothesis that a stereotype's cumulative self-fulfilling effect can influence how targets are viewed by others. Perceivers judged a target according to the target's behavior, inferring that a target who confirmed the stereotype of women generally had stereotypically feminine traits and abilities, and that a target who confirmed the stereotype of men generally had stereotypically masculine traits and abilities. Importantly, perceivers made these correspondent inferences even when they were privy to the situational affordances that gave rise to the target's behavior. Although one might wonder whether this tendency reflects a failure of perceivers to notice the situational affordances more than a failure to take them into account, the well-established tendency for people to make correspondent inferences about others whom they know lacked free choice suggests that perceivers fell prey to the fundamental attribution error when judging the target.

We also want to emphasize that a failure of perceivers to take the situational affordances into account is not merely a replication of the fundamental attribution error. Because the experiment focused exclusively on behaviors that had been elicited by a stereotype's prior

cumulative self-fulfilling effect, the results highlight how the fundamental attribution error can serve as a mechanism whereby a stereotype's cumulative self-fulfilling effect can perpetuate from one social interaction to another. Indeed, the stereotypic judgments that perceivers made were not caused by a reliance on sex stereotypes, but instead by a reliance on the target's behavior – behavior that confirmed sex stereotypes to varying degrees because of the stereotype's prior cumulative self-fulfilling effect. These results are particularly troubling because they applied not only to judgments of sex-typed traits, but also to judgments of general academic competence, career aptitude, and career suitability; judgements that could encourage perceivers to exclude or wrongfully channel some targets into areas of study and work on the basis of the accumulating effects of stereotypic beliefs.

General Discussion

The idea that stereotypes can have cumulative self-fulfilling effects regularly appears in the psychological literature (Jussim, 2012). Such assertions, though theoretically compelling, have been purely speculative, made in the absence of empirical support. The present research provided an initial test of the hypothesized effect, and strongly supported the accumulation hypothesis. Across the board, targets confirmed stereotypes about weight and sex to a greater extent when two perceivers held stereotypic expectations about them than when only one did. In addition, targets' confirmatory behavior had downstream consequences. Naïve perceivers who had not caused targets to confirm the stereotypes nonetheless judged them according to the stereotypic behaviors they had been channeled to adopt, an effect that even occurred among perceivers who were aware that targets' behavior was shaped by the actions of others.

These findings advance theoretical and empirical understanding about the self-fulfilling nature of stereotypes in three critical respects. First, they help to clarify how stereotype-based

self-fulfilling prophecies can create or exacerbate social problems despite that false expectations typically have only modest self-fulfilling effects on targets' behavior within dyadic relations. Second, they show that situational affordances, which have long been thought to contribute to the self-fulfilling effect of stereotypes, can also operate as a mechanism of accumulation. Because this research demonstrated the effect of this mechanism among independent perceivers, it also explains how multiple perceivers who do not know one another can, through their collective actions, construct a situation that more strongly encourages targets to confirm a stereotype than can the actions of individual perceivers acting singly. Third, the findings suggest that the stereotypic behaviors that targets exhibit as a result of a stereotype's cumulative self-fulfilling effect can become the foundation upon which new perceivers form impressions of them. Thus, the cumulative self-fulfilling effects of stereotypes may create seemingly defensible positions from which new perceivers may afford targets opportunities and constraints that could widen the gap between advantaged and disadvantaged groups.

Self-Fulfilling Prophecies and Social Problems

For most of the 20th century, the psychological literature characterized the self-fulfilling prophecy as a powerful phenomenon capable of producing large-scale social problems in areas such as hiring, education, wages, and health care (Merton, 1948; Ross, Lepper, & Ward, 2010; Snyder, 1984). Such claims were appealing because they offered a psychological and sociological explanation for undeniable social injustice. Yet, as empirical evidence mounted, it seemed that claims of powerful self-fulfilling prophecies may have been overstated. Empirical research consistently showed that self-fulfilling prophecy effects were relatively modest in magnitude (Rosenthal, 1994, 2003). Even so, such findings did not preclude the possibility that powerful self-fulfilling prophecy effects could still arise through processes of accumulation.

In support of this possibility, Madon and colleagues showed that the self-fulfilling effects of parents' expectations on their adolescents' alcohol use accumulate across mothers and fathers (Madon et al., 2004), and over time among particular adolescents (Madon, Willard, Guyll, Trudeau & Spoth, 2006). However, because their findings pertained only to interpersonal expectations, which are idiosyncratic and, therefore, an unlikely cause of widespread social injustice, it was not clear how the accumulation processes they identified could produce the large-scale social problems proposed by the theoretical literature. Consideration of this issue naturally suggested the influence of stereotypes, which have always been at the heart of claims linking self-fulfilling prophecies to social problems. Thus, despite initial empirical support for general processes of accumulation, an important question remained unanswered: Can the self-fulfilling effects of stereotypes accumulate? The results of the present research indicated that they can: Consistently, the stereotypic expectations of multiple perceivers caused targets to more strongly confirm a stereotype than did the stereotypic expectations of individual perceivers.

Magnitude of Stereotype-Based Cumulative Self-fulfilling Prophecy Effects

Because the cumulative self-fulfilling effect of stereotypes is fundamentally an issue of magnitude, it is useful to compare the effects reported in this research to those reported in the broader literature. In this research, the average self-fulfilling prophecy effect associated with a single perceiver's stereotypic expectation (d = .65) was similar to the meta-analytic effect size of interpersonal self-fulfilling prophecy effects within dyadic relations (d = .60; Rosenthal, 1994). The present research also indicated, however, that a second perceiver's stereotypic expectation had an additional self-fulfilling effect beyond that associated with the first perceiver's self-fulfilling effect. Across the experiments, the self-fulfilling effect uniquely attributable to a second perceiver's stereotypic expectation was d = 1.04, on average. Furthermore, because of

this additional unique effect, the average total self-fulfilling effect attributable to both perceivers combined was d = 1.69, (i.e., .65 + 1.04 = 1.69), which is more than twice the magnitude of the self-fulfilling effect elicited by one perceiver in this research and in the broader literature.

The accumulation effect observed herein is important because it suggests that the typical self-fulfilling effects reported in the literature may underestimate the power of stereotypes. This is because the literature has focused exclusively on the self-fulfilling effect of stereotypes within dyadic relations even though stereotypes are often consensual. Thus, different perceivers may hold similar stereotypic expectations about a target. If these expectations are inaccurate, then each perceiver may exert a self-fulfilling effect that could combine with the self-fulfilling effects of other perceivers to ultimately have a powerful influence on the target's behavior. Such a process is troublesome because it may exacerbate social problems by virtue of creating unjust social trajectories. Moreover, if new perceivers fail to adjust for the fact that a stereotype's prior self-fulfilling effect caused a target's stereotypic behavior, as was the case in this research, then these social trajectories could become self-sustaining, or at the very least difficult to change.

Situational Affordances

The present research tested for the accumulation of stereotype-based self-fulfilling prophecy effects in the context of situational affordances. When perceivers construct situational affordances in accord with their stereotypic expectations, they encourage targets to behaviorally confirm a stereotype by channeling their behavior in the direction of the stereotype. The potential for a stereotype's self-fulfilling effect to accumulate across perceivers arises when multiple perceivers provide similar situational affordances to the same target. Consistent with this mechanism, the targets in this research more strongly confirmed stereotypes about weight and sex when multiple perceivers had independently provided them with a situational affordance that

encouraged confirmation of the stereotypes than when only one perceiver had done so

Importantly, however, targets were not compelled to act in stereotypic ways. Although the situational affordances channeled their behavior in the direction of the stereotypes, all targets had the freedom to predominantly disconfirm the stereotypes. The fact that they did not exercise this freedom suggests that more than just the objective features of the situational affordances impinged on their behavior. The established propensity for people to actively construct social reality suggests that targets may have confirmed the stereotypes because of the way they interpreted the situational affordances. For example, we speculated that the targets in Experiment 1 may have perceived a large amount of candy as a form of cover that permitted them to take many pieces without detection, and that the targets in Experiment 3 may have considered the interpersonal consequences of their article selections for the anticipated meeting with their partners. One way to understand the results of this research, therefore, is that perceivers paved a behavioral path for targets that targets followed because of the way they construed it.

This explanation also suggests that different circumstances might have led targets to construe the same situational affordances in ways that discouraged them from confirming the stereotypes. For example, the literature on dyadic self-fulfilling prophecies indicates that targets are less susceptible to self-fulfilling prophecies when they are aware of perceivers' expectations (Hilton & Darley, 1985). Applying this finding to the present research suggests that targets might have tended to disconfirm the stereotypes had they attributed the situational affordances to perceivers' stereotypic expectations. For instance, if targets believed that they received a large amount of candy because of their perceived weight, or predominantly sex-typed articles because of their perceived sex, then they may have intentionally taken less candy or fewer sex-typed articles. Thus, the way that targets interpret a situation may have a bearing on a stereotype's

cumulative self-fulfilling effect.

Other Potential Mechanisms of Accumulation

The present focus on situational affordances as a mechanism of a stereotype's cumulative self-fulfilling effect is consistent with historical analyses of stereotype-based self-fulfilling prophecies. However, consideration of the broader social psychological literature suggests other potential mechanisms of the accumulation process. We next discuss two that seem particularly likely in light of the empirical evidence: self-concept change and amplification.

Self-concept change. Perceivers' expectations can change targets' self-concepts through an iterative interaction sequence (Darley & Fazio, 1980). Though typically observed within dyadic relations, such changes could be greater when multiple perceivers hold similar expectations about the same target due to the perceived credibility ascribed to consensual beliefs. Analogous to Kelly's covariation model (1973), when an expectation is idiosyncratic to one perceiver, targets may conclude that it says more about the perceiver's characteristics than their own, thus weakening the expectation's apparent credibility and reducing self-concept change.

By contrast, when an expectation is consensually shared by multiple perceivers, as in the case of some stereotypes, it may be given more credibility. Targets may wonder, quite reasonably, why multiple perceivers would expect them to possess or lack a particular characteristic if it were not true? This could lead targets to conclude that the expectation says more about their own characteristics than those of the perceivers, thus strengthening the expectation's apparent credibility and increasing self-concept change. Although the number of perceivers who must share an expectation for this to happen is an empirical question, it is possible that two perceivers might be enough, in which case the critical factor driving the effect may be a categorical shift from an idiosyncratic expectation to one that is consensual.

However, because self-concept change is not behavioral, it does not satisfy the self-fulfilling prophecy criterion of behavioral confirmation. Nevertheless, changes to targets' self-concepts could initiate behavioral confirmation through self-verification strivings. Targets who internalize a perceiver's false expectation may engage in self-verification processes that subsequently cause them to behaviorally confirm the false expectation (Madon et al., 2008; Scherr et al., 2011; Snyder & Swann, 1978). The credibility ascribed to consensually shared stereotypic expectations may exacerbate this effect. This is because consensually shared expectations may produce greater changes to targets' self-concepts than idiosyncratic expectations, and correspondingly, elicit more stereotype-confirming behavior from targets via self-verification processes. Moreover, to the extent that targets discount a single perceiver's stereotypic expectation, a target's susceptibility to a stereotype's self-fulfilling effect may require an emergent consensus of expectations. That is to say, there may be circumstances under which accumulation is necessary for any self-fulfilling prophecy effect to occur, and in such cases synergistic accumulation may be especially likely, such as what happened in Experiment 1.

Amplification. In this research, the self-fulfilling effect of stereotypes accumulated across perceivers who never interacted with one another, an effect that occurred because multiple perceivers independently provided a target with a similar situational affordance. This result shows that the accumulation process does not require social interaction among perceivers. However, when perceivers do interact, there their expectations may become amplified. For example, Willard et al. (2012) showed that perceivers' false expectations about a target's hostility became more extreme when perceivers interacted with another perceiver who also believed the target had a hostile personality.

The tendency for social interaction to amplify perceivers' expectations has important

implications for a stereotype's cumulative self-fulfilling effect. First, the more amplified perceivers' expectations, the more likely perceivers may be to construct situational affordances that encourage confirmation of a stereotype. Second, an amplified expectation may change targets' self-concepts more than an expectation that has not been amplified, thus also evoking more stereotype-consistent behavior from targets through the self-verification processes described above. Examination of these amplification processes awaits research paradigms that permit multiple perceivers to interact prior to treating targets in stereotypic ways.

Limitations

External validity. The methods of this research created an artificial situation in which perceivers independently made behavioral decisions (e.g., selecting candy for a target) on the basis of targets' social group membership (e.g., heavy versus thin) that were passed along to targets through experimental procedures. The experimental control created by this approach was useful for testing whether the accumulation of self-fulfilling prophecy effects across perceivers is possible, but it also simplified the accumulation process. Notably, it removed the potential influence of a host of factors that might typically be present in the naturalistic environment (e.g., certainty of expectations, perceiver and target relations, self-verification strivings, social interaction, etc.), and which could influence a stereotype's cumulative self-fulfilling effect.

Although this limitation may raise questions about the external validity of the findings, it also highlights a potential contribution of the present research inasmuch as it suggests that stereotype-based self-fulfilling prophecy effects may be especially likely to accumulate when targets' opportunities and constraints are tightly controlled by perceivers; historical and contemporary examples include American segregation, South African apartheid, caste systems, and cultures that oppress women. We are not suggesting that the accumulation of stereotype-

based self-fulfilling prophecies is necessarily limited to these kinds of contexts, but rather that these contexts may be especially conducive to eliciting the effect. Indeed, Madon et al.'s (2004) finding that self-fulfilling prophecy effects accumulated across perceivers in the context of parents' expectations about their adolescents' future alcohol use suggests that the accumulation of self-fulfilling prophecy effects may reflect a general process of interpersonal influence that can occur in less restrictive situations and contexts than noted here.

Patterns of accumulation. The present research provided clear and consistent evidence of accumulating self-fulfilling prophecy effects, but the pattern of these effects differed across the stereotypes examined. Whereas synergistic accumulation characterized the overweight stereotype and the stereotype of men, concurrent accumulation characterized the stereotype of women. Several points regarding these differences deserve mention. First, we cannot easily explain why different accumulation patterns emerged for the stereotypes of women and men. Initially, we considered the possibility that the women in our sample might have felt more comfortable selecting stereotypically masculine articles than the men did selecting stereotypically feminine articles. However, there was no evidence to support this interpretation as women and men responded similarly to the article sets. We also cannot attribute the difference to method variance because the same method was used to test the cumulative effects of both stereotypes. The absence of any clear explanation points to the need for future research.

Second, the tendency for synergistic accumulation to predominate the findings may seem to suggest that it is the more prevalent of the two patterns, especially because it also characterized Madon et al.'s (2004) examination of interpersonal cumulative self-fulfilling prophecy effects. Although possible, an alternative interpretation is that synergistic accumulation is particularly likely to emerge with the kinds of behaviors that have been used to test the

accumulation hypothesis. Both this research and Madon et al., (2004) examined behaviors that were not bound by a target's ability. The amount of candy a target takes, the type of articles a target selects, and the amount of alcohol an adolescent drinks do not require any special talent, training, or aptitude that could restrict targets from confirming a false expectation. As a result, such behaviors may be particularly susceptible to the accelerating pattern of synergistic accumulation. Other behaviors, such intelligence or athleticism, may be less susceptible to this accelerating pattern because targets' true abilities may set boundaries that are difficult to surpass. In such cases, concurrent accumulation may be more typical. Further, when accumulation does occur with ability-based behaviors, accumulation may be maximized with only a few perceivers because targets may quickly reach the limit of their true abilities.

Finally, ample research in social psychology indicates that motivational, cognitive, and dispositional factors can moderate the strength of self-fulfilling prophecy effects (Snyder & Stukas, 1999). An important next step in theory building is to investigate whether these factors also affect the presence of concurrent and synergistic accumulation. It is possible, for example, that synergistic accumulation is rendered more likely by conditions that intensify self-fulfilling prophecy effects, such as when perceivers are motivated to control targets' behavior, targets have lower status or power than perceivers, and perceivers' expectations are clear but targets' self-views are unclear. Conversely, concurrent accumulation may be rendered more likely by conditions that mitigate self-fulfilling prophecy effects, such as when perceivers are motivated to get along with targets, targets have higher status or power than perceivers, and perceivers' expectations are unclear but targets' self-views are clear. Discovering the motivational, cognitive, and dispositional factors that tend to elicit one or the other pattern of accumulation will provide greater theoretical understanding about the accumulation process.

Demand characteristics. A key finding of this research was that perceivers constructed situational affordances that proximally caused targets to confirm the stereotypes. We speculated that this effect occurred because of the way targets construed the situation. However, one might wonder whether demand characteristics offers a more parsimonious explanation (e.g., *"they gave me a lot of candy, they must want me to take many pieces."*). Two considerations argue against this interpretation. First, the suspicion rates were very low and, second, in the one instance in which removing the data of suspicious participants affected the results (Experiment 1), it led to a stronger, not weaker, accumulation effect, suggesting the influence of psychological reactance (Brehm, 1966) rather than demand characteristics.

Number of perceivers. Because the present research examined accumulation in the context of two perceivers, it cannot address what happens when more than two perceivers are involved. The theorized link between self-fulfilling prophecies and social problems implies that a stereotype's cumulative self-fulfilling effect strengthens with additional perceivers. However, it is also possible that two is special, meaning that a stereotype's cumulative self-fulfilling effect may level-off after the minimum requirement of two perceivers has been reached. The more general issue is that there may sometimes be a curvilinear, asymptotic relationship between the number of perceivers who hold a stereotypic expectation about a target and the target's tendency to behaviorally confirm the stereotype. Whether this is the case will require future research to test the accumulation hypothesis across more than two perceivers. If future research supports the "two is special" hypothesis, then long-standing assumptions about the way that self-fulfilling prophecies contribute to social problems may need to be reconsidered.

Protected status. Finally, even though our results showed that stereotypes can have cumulative self-fulfilling effects, we demonstrated this effect with social groups about whom

people feel relatively comfortable expressing negative attitudes (e.g., Haines et al., 2016; Leskinen, Rabelo, & Cortina, 2015; Puhl & Heuer, 2009). As such, our procedures might have increased the chances that participants would treat targets in line with their stereotypes, a necessary step in the self-fulfilling prophecy process. This raises the possibility that perceivers may be less inclined to apply their stereotypes to groups with greater protected status (Madon, Smith, & Guyll, 2005), such as those associated with race or military service, in which case a stereotype's cumulative self-fulfilling effect may be mitigated or even preempted.

Conclusion

Self-fulfilling prophecies are hypothesized to contribute to social problems by generating and perpetuating group inequalities. The cumulative self-fulfilling effect of stereotypes is one means through which such inequalities can arise. Consistent with this process, the present research provided the first empirical evidence that the self-fulfilling effect of stereotypes can accumulate across perceivers. Targets more strongly confirmed stereotypes about weight and sex when two perceivers treated them in stereotypic ways than when only one did. Moreover, targets’ confirmatory behavior biased new perceivers' judgments of them, thereby showing how the cumulative self-fulfilling effects of stereotypes can transcended beyond the original interactions that produced them. These findings are consistent with a long line of research within social psychology emphasizing the power of beliefs to create reality and demonstrate how even small self-fulfilling prophecy effects can contribute to social problems via an accumulation process.

References

Brehm, J. W. (1966). *A theory of psychological reactance*. Oxford, England: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Crandall, C. S. (1994). Prejudice against fat people: Ideology and self-interest. *Journal of Personality and Social Psychology, 66*(5), 882-894. doi:10.1037/0022-3514.66.5.882

Crandall, C. S., & Eshleman, A. (2003). A justification-suppression model of the expression and experience of prejudice. *Psychological Bulletin, 129*, 414-446. doi: 10.1037/0033-2909.129.3.414

Darley, J. M., & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist, 35*, 867-881. doi:10.1037/0003-066X.35.10.867

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology* (Vol. 23, pp. 75-109). Academic Press, San Diego, CA, 1990.

Fiske, S. T. (2004). *Social beings: A core motives approach to social psychology*. Hoboken, NJ: John Wiley & Sons.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Haines, E. L., Deaux, K., & Lofaro, N. (March, 2016). The times they are a-changing…or are they not? A comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly, 9*, 1-11. doi:10.1177/0361684316634081

Hilton, J. L., & Darley, J. M. (1985). Constructing other persons: A limit on the effect. *Journal of Experimental Social Psychology, 21*, 1-18. doi: 10.1016/0022-1031(85)90002-2

Jones, E. E., & Davis, K. E. (1965). A theory of correspondent inferences: From acts to

dispositions. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219 – 266). New York, NY: Academic Press.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3*, 1-24. doi:10.1016/0022-1031(67)90034-0

Jussim, L. (2012). *Social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy.* Oxford University Press, New York, NY. doi:10.1093/acprof:oso/9780195366600.001.0001

Jussim, L., Eccles, J. S., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, pp. 281-388). San Diego, CA: Academic Press. doi:10.1016/S0065-2601(08)60240-3

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist, 28*, 107-128. doi: 10.1037/h0034225

Klein, O., & Snyder, M. (2003). Stereotypes and behavioral confirmation: From interpersonal to intergroup perspectives. *Advances in experimental social psychology, vol. 35.* (pp. 153-234) Elsevier Academic Press, San Diego, CA. doi:10.1016/S0065-2601(03)01003-7

Koffka, K. (1935). *The principles of gestalt psychology.* Harcourt, Brace and Company: NY.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological review*, *103*, 284-308.

Leskinen, E. A., Rabelo, V. C., & Cortina, L. M. (2015). Gender stereotyping and harassment: A "catch-22" for women in the workplace. *Psychology, Public Policy, and Law*, *21*, 192-204.

MacKinnon, D. P., Cheong, J., & Pirlott, A. G. (2012). Statistical mediation analysis. *APA handbook of research methods in psychology* (Vol. 2, pp. 313-331). American Psychological

Association, Washington, DC. doi:10.1037/13620-018

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593-614. doi:10.1146/annurev.psych.58.110405.085542

Madon, S., Guyll, M., Aboufadel, K., Montiel, E., Smith, A., Palumbo, P., & Jussim, L. (2001). Ethnic and national stereotypes: The Princeton trilogy revisited and revised. *Personality and Social Psychology Bulletin, 27*, 996-1010. doi:10.1177/0146167201278007

Madon, S., Guyll, M., Buller, A. A., Scherr, K. C., Willard, J., & Spoth, R. (2008). The mediation of mothers self-fulfilling effects on their children's alcohol use: Self-verification, informational conformity, and modelling processes. *Journal of Personality and Social Psychology, 95*, 369-384. doi: 10.1037/0022-3514.95.2.369

Madon, S., Guyll, M., Hilbert, S. J., Kyriakatos, E., & Vogel, D. L. (2006), Stereotyping the stereotypic: When individuals match social stereotypes. *Journal of Applied Social Psychology, 36*, 178–205. doi:10.1111/j.0021-9029.2006.00057.x

Madon, S., Guyll, M., Spoth, R., & Willard, J. (2004). Self-fulfilling prophecies: The synergistic accumulation of parents' beliefs on children's drinking behavior. *Psychological Science 15*, 837-845. doi:10.1111/j.0956-7976.2004.00764.x

Madon, S., Smith, A. E., & Guyll, M. (2005). Social norms regarding protected status and threat reactions to the stigmatized. *Journal of Applied Social Psychology*, 35, 572-602. doi: 10.1111/j.1559-1816.2005.tb02136.x

Merton, R. K. (1948). The self-fulfilling prophecy. *Antioch Review, 8,* 193-210. doi:10.2307/ 4609267

Puhl, R. & Brownell, K. D. (2001). Bias, discrimination, and obesity. *Obesity Research, 9,* 788-805. doi:10.1038/oby.2001.108

Puhl, R. M., & Heuer, C. A. (2009). The stigma of obesity: A review and update. *Obesity, 17,* 941-964. Doi: 10.1038/oby.2008.636

Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (2001). *Applied regression analysis: A research tool* (2nd Edition). Springer-Verlag New York, Inc.: New York, NY.

Rist, R. C. (2000). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review, 70,* 266-301. doi: 10.17763/haer.70.3.1k0624l6102u2725

Rosenthal, R. (1994). Interpersonal expectancy effects: A 30-year perspective. *Current Directions in Psychological Science, 3,* 176-179. doi: 10.1111/1467-8721.ep10770698

Rosenthal, R. (2003). Covert communication in laboratories, classrooms, and the truly real world. *Current Directions in Psychological Science, 12,* 151-154. doi:10.1111/1467-8721.t01-1-01250

Rosenthal, R., & Jacobson, L. (1968). Teacher expectations for the disadvantaged. *Scientific American, 218,* 19-23.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173 – 220). New York, NY: Academic Press. doi:10.1016/S0065-2601(08)60357-3

Ross, L., Lepper, M., & Ward, A. (2010). History of social psychology: Insights, challenges, and contributions to theory and application. *Handbook of social psychology* (Vol. 1, pp. 3-50) John Wiley & Sons Inc., Hoboken, NJ. doi:10.1002/9780470561119.socpsy001001

Scherr, K. C., Madon, S., Guyll, M., Willard, J., & Spoth, R. (2011). Self-verification as a mediator of mothers' self-fulfilling effects on adolescents' educational attainment. *Personality and Social Psychology Bulletin, 37,* 587-600. doi: 10.1177/0146167211399777

Snyder, M. (1984). When belief creates reality. In M. P. Zanna (Ed.), *Advances in experimental psychology* (Vol. 18, pp. 247-305). Orlando, FL: Academic Press. doi:10.1016/S0065-2601(08)60146-X

Snyder, M., & Swann, W. B. (1978). Behavioral confirmation in social interaction: From social perception to social reality. *Journal of Experimental Social Psychology, 14,* 148-162.

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35,* 656-666. doi:10.1037/0022-3514.35.9.656

Snyder, M., & Stukas, A. A., Jr. (1999). Interpersonal processes: The interplay of cognitive, motivational, and behavioral activities in social interaction. *Annual Review of Psychology, 50,* 273-303. doi: 10.1146/annurev.psych.50.1.273

Steiger, J. H. (2004). Beyond the *F* test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods, 9,* 164–182. doi:10.1037/1082-989X.9.2.164

Willard, J., Madon, S., Guyll, M., Scherr, K. C., & Buller, A. A. (2012). The accumulating effects of shared expectations. *European Journal of Social Psychology, 42,* 497.

Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology, 10,* 109-120. doi:10.1016/0022-1031(74)90059-6

Wuensch, K. L.(2012). Using SPSS to obtain a confidence interval for Cohen's d. Retrieved from http://core.ecu.edu/psyc/wuenschk/SPSS/CI-d-SPSS.pdf.

Table 1

Experiment 1: Perceivers' Trait Judgments of the Target

	Expectation Condition						
	Heavy		Thin		Mean Difference	Effect Size	
Trait Judgment	*M*	*SD*	*M*	*SD*	95% CI	*d*	95% CI
Willpower	3.56	1.08	3.95	1.09	.20, .59	0.36	.18, .54
Self-Control	4.18	1.20	4.67	1.08	.29, .70	0.43	.25, .61
Outgoing	3.24	1.30	3.19	1.22	-.27, .18	0.04	-.14, .21
Intelligent	4.67	0.81	4.79	0.93	-.04, .27	0.13	-.05, .31
Religious	4.39	0.95	4.34	1.15	-.23, .14	0.04	-.14, .22

Note: The *df* was 479 for willpower, self-control, outgoing, and intelligent, and 478 for religious. *M* = mean, *SD* = standard deviation, *d* = Cohen's *d.* Cohen's *d* and the corresponding confidence intervals were calculated with a script developed by Wuensch (2012).

Table 2

Experiment 2: Average Number of Stereotypically Feminine, Gender-Neutral, and Stereotypically Masculine Articles Perceivers Selected for the Target.

Expectation Condition	Article Type Selected								
	Stereotypically Feminine			Gender-Neutral			Stereotypically Masculine		
	M	Mode	*SD*	*M*	Mode	*SD*	*M*	Mode	*SD*
Feminine	2.00	2	0.96	0.64	0	0.73	0.36	0	0.66
Gender-Neutral	0.68	0	0.84	1.79	2	0.84	0.53	0	0.69
Masculine	0.07	0	0.26	0.86	0	0.86	2.07	2	0.83

Note: M = mean, *SD* = standard deviation.

Table 3

Experiment 3: Average Number of Stereotypically Feminine, Gender-Neutral, and Stereotypically Masculine Articles Targets Selected to Read.

	Article Type Selected					
	Stereotypically Feminine		Gender-Neutral		Stereotypically Masculine	
Experimental Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Double-Feminine	1.81	0.69	1.15	0.72	0.00	0.00
Single-Feminine	1.40	0.71	1.54	0.76	0.00	0.00
Gender-Neutral	0.60	0.50	1.88	0.53	0.52	0.51
Single-Masculine	0.00	0.00	1.95	0.80	0.95	0.69
Double-Masculine	0.00	0.00	0.77	0.65	2.20	0.65

Note: M = mean, *SD* = standard deviation.

Table 4

Experiment 4: Perceivers' Judgments of the Target.

	Personality Traits		General Academic Competence		Aptitude in Math and Science		Career Suitability	
Target Behavior	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Strong-feminine	4.60	0.67	4.43	0.81	3.82	0.62	5.81	0.84
Moderate-feminine	4.16	0.71	4.61	0.70	4.14	0.69	4.92	1.20
Gender-neutral	4.05	0.56	4.93	0.80	4.13	0.72	4.59	1.07
Moderate-masculine	3.66	0.43	5.07	0.82	4.86	0.79	3.24	0.88
Strong-masculine	3.70	0.54	5.59	0.92	5.52	1.00	2.87	1.07

Note: M = mean, *SD* = standard deviation.

Supplemental Table 1

Experiment 1: Phase 1 Descriptive Statistics.

Variables		(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Expectation[a]		-----	-.18**	-.21**	.02	-.07	.02	.52**
(2) Willpower[b]			-----	.41**	.28**	.30**	.02	-.09*
(3) Self-control[b]				-----	.12**	.21**	.15**	-.15**
(4) Outgoing[b]					-----	.02	-.15*	.05
(5) Intelligent[b]						-----	.18**	-.10*
(6) Religious[b]							-----	-.02
(7) Bin selected[cd]								-----
	M	0.50	3.75	4.43	3.22	4.73	4.37	2.55
	SD	0.50	1.10	1.17	1.26	0.87	1.05	0.79
	N	482	481	481	481	481	480	482

Note. M = mean, *SD* = standard deviation. *N* = number of perceivers. [a]The expectation was coded as 0 (non-overweight expectation) and 1 (overweight expectation). [b]Willpower, self-control, outgoing, intelligent, and religious refer to perceivers' trait judgments of the target. [c]Bin selected corresponds to the bin of candy that perceivers selected for the target to receive, coded as 1 (2 pieces of candy), 2 (4 pieces of candy), 3 (20 pieces of candy), and 4 (40 pieces of candy). [d]Correlations computed with Spearman's rank-order correlation.

Supplementary Table 2

Experiment 1: Phase 2 Descriptive Statistics.

Variables		(1)	(2)	(3)
(1) Expectation[a]		-----	.57**	.18**
(2) Candy received[b]			-----	.28**
(3) Candy taken[c]				-----
	M	1.00	28.59	3.61
	SD	0.83	17.78	4.03
	N	241	241	241

Note. *M* = mean, *SD* = standard deviation, *N* = number of triads. [a]The expectation was coded as 0 (no-overweight expectation), 1 (single-overweight expectation), and 2 (double-overweight expectation). [b]Candy received refers to the average amount of candy given to targets in Phase 2. [c]Candy taken refers to the average amount of candy taken by targets.

Supplementary Table 3

Experiment 2: Descriptive Statistics.

Variables	(1)	(2)	(3)	(4)	(5)
(1) Expectation[a]	-----	-.73**	.09	.67**	-.63**
(2) Feminine article selections[b]		-----	-.47**	-.62**	.51**
(3) Gender-neutral article selections[b]			-----	-.40**	-.03
(4) Masculine article selections[b]				-----	-.50**
(5) Sex-typed trait judgments[c]					-----
M	2.01	0.92	1.07	1.01	7.94
SD	0.83	1.11	0.94	1.07	2.68
N	123	123	123	123	122

Note. *M* = mean, *SD* = standard deviation. *N* = number of perceivers. [a]The expectation was coded as 1 (feminine expectation), 2 (gender-neutral expectation), and 3 (masculine expectation). [b]The feminine, gender-neutral, and masculine article selections refer to the number of stereotypically feminine, gender-neutral, and stereotypically masculine articles that participants selected for the target to read. [c]Sex-typed trait judgments refers to participants' judgments of the target's femininity, with higher scores indicating greater perceived femininity.

Supplementary Table 4

Experiment 3: Descriptive Statistics.

Variables	(1)	(2)	(3)	(4)
(1) Experimental manipulation[a]	-----	.80**	.08	-.82**
(2) Feminine article selections[b]		-----	-.36**	-.64
(3) Gender-neutral article selections[b]			-----	-.49
(4) Masculine article selections[b]				-----
M	3.06	0.80	1.43	0.72
SD	1.44	0.89	0.82	0.95

Note. *M* = mean, *SD* = standard deviation. [a]The experimental manipulation was coded as 1 (double-masculine), 2 (single-masculine), 3 (gender-neutral), 4 (single-feminine) and 5 (double-feminine). [b]The feminine, gender-neutral, and masculine article selections refer to the number of stereotypically feminine, stereotypically masculine, and gender-neutral articles selected by targets, respectively.

Supplementary Table 5

Experiment 4: Descriptive Statistics.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Target behavior[a]	-----	.00	.48**	-.44**	-.59**	.72**	-.08	-.44**
(2) Situational awareness[b]		-----	.03	.03	-.00	-.02	-.10	-.09
(3) Sex-typed traits[c]			-----	-.46**	-.51**	.66**	.08	.14*
(4) General academic competency[c]				-----	.70**	-.48**	.02	-.27**
(5) Aptitude in math and science[c]					-----	-.73**	.00	-.34**
(6) Suitability for sex-typed careers[c]						-----	.05	.32**
(7) Article titles correctly recalled							-----	.06
(8) Memory for article summaries								-----
M	3.00	0.50	4.04	4.93	4.49	4.28	2.83	5.62
SD	1.42	0.50	0.68	0.90	0.99	1.49	0.64	0.63
N	230	230	225	226	228	227	230	230

Note. *M* = mean, *SD* = standard deviation. *N* = number of perceivers. [a]Target behavior was coded as 1 (strong-masculine), 2 (moderate-masculine), 3 (gender-neutral), 4 (moderate-feminine), and 5 (strong-feminine). [b]Situational awareness was coded as 0 (unaware) and 1 (aware). [c]Sex-typed traits, general academic competency, aptitude in math and science, and suitability for sex-typed careers refer to perceivers' judgments of the target. Higher values on these judgments correspond to greater feminine sex-typed traits, greater general academic competency, greater aptitude in math and science, and better-suited for stereotypically feminine careers.

Supplementary Table 6

Experiment 4: ANOVA Results.

Sex-Typed Traits

Source	*df*	*F*	*p*	η_p^2	95% CI
Target behavior	4	19.40	<.000	.27	.17, .33
Constraint awareness		0.39	.533	.00	.00, .02
Target behavior × Constraint awareness		1.04	.386	.02	.00, .04
Error	215				
Total	225				

General Academic Competency

Source	*df*	*F*	*p*	η_p^2	95% CI
Target behavior	4	13.95	<.000	.21	.12, .27
Constraint awareness	1	0.23	.636	.00	.00, .00
Target behavior × Constraint awareness	4	1.59	.178	.03	.00, .06
Error	216				
Total	226				

Aptitude for Math and Science

Source	*df*	*F*	*p*	η_p^2	95% CI
Target behavior	4	36.24	<.000	.40	.31, .46
Constraint awareness	1	0.01	.909	.00	.00, .00
Target behavior × Constraint awareness	4	1.29	.275	.02	.00, .05
Error	218				
Total	228				

Career Suitability

Source	*df*	*F*	*p*	η_p^2	95% CI
Target behavior	4	63.62	<.000	.54	.46, .59
Constraint awareness	1	0.20	.656	.00	.00, .01
Target behavior × Constraint awareness	4	1.08	.366	.02	.00, .04
Error	217				
Total	227				

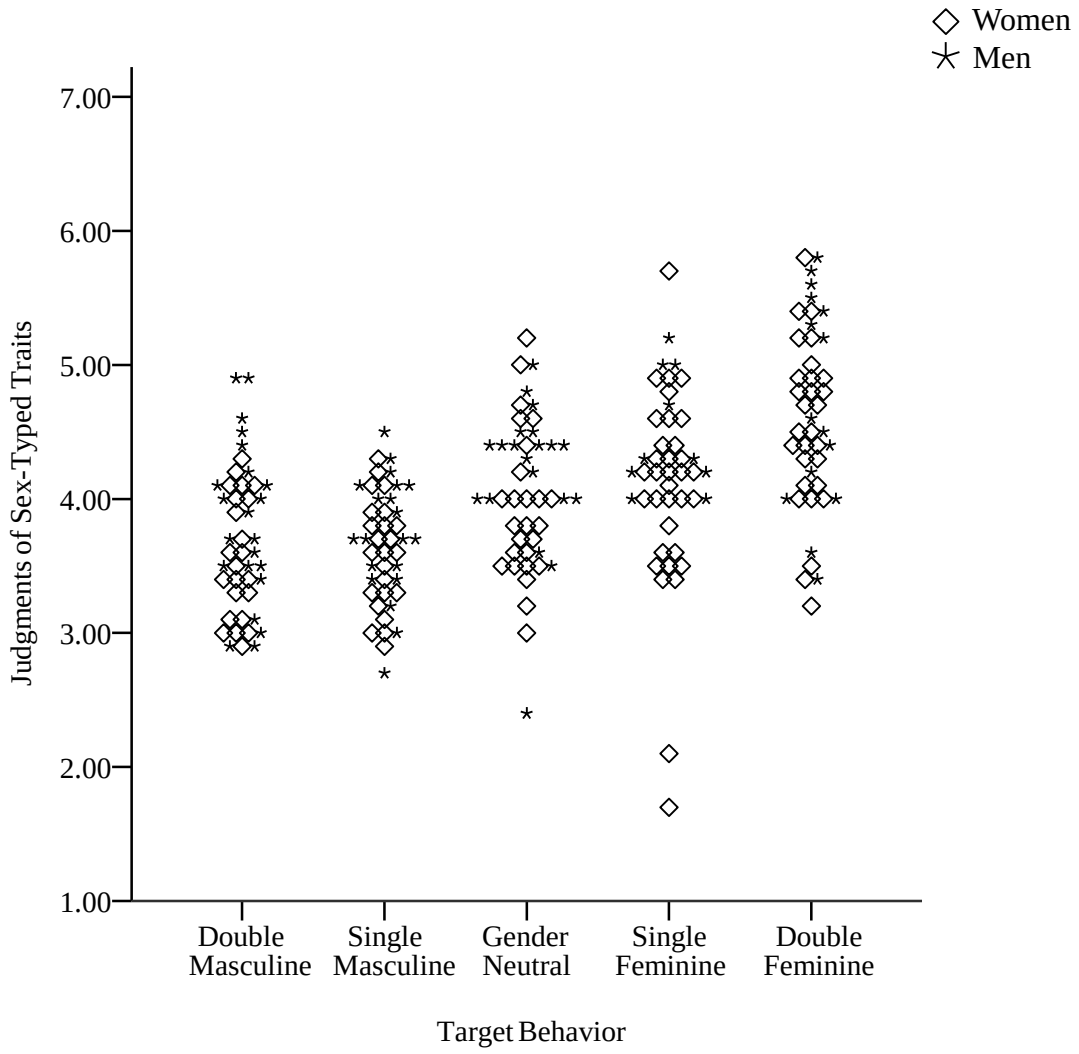
Note: *df* = degrees of freedom.

Supplementary Table 7

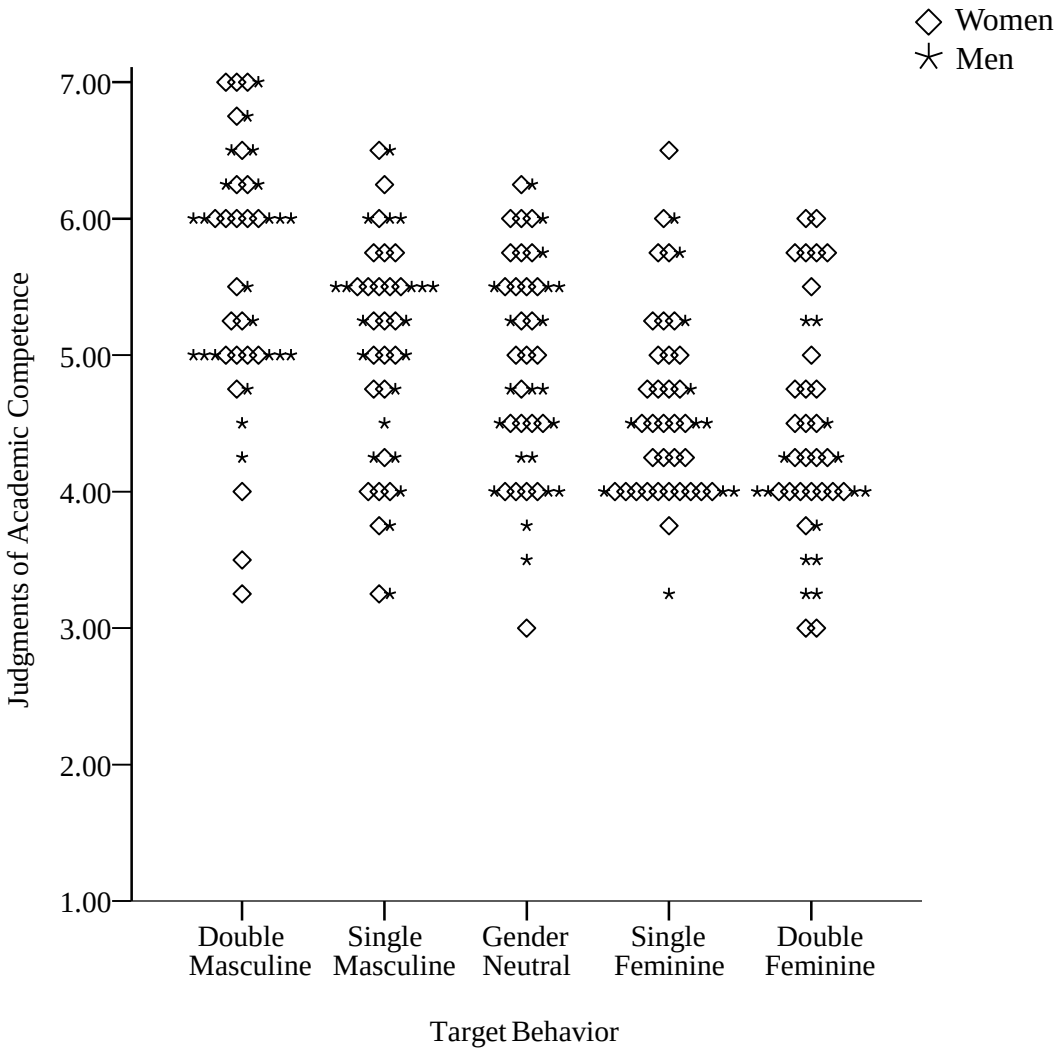
Experiment 4: Pairwise Comparisons of Target Behavior.

Comparison	*t*	*p*	*d*	95% CI
Personality Traits (*df* = 220)				
Double-feminine vs. Single feminine	3.84	.001	0.81	0.39, 1.23
Double-feminine vs. Gender-neutral	4.75	<.001	0.99	0.57, 1.41
Double-feminine vs. Single masculine	7.83	<.001	1.65	1.21, 2.09
Double-feminine vs. Double masculine	7.53	<.001	1.58	1.14, 2.01
Single-feminine vs. Gender-neutral	0.86	.911	0.18	-0.23, 0.59
Single-feminine vs. Single-masculine	3.99	.001	0.84	0.42, 1.26
Single-feminine vs. Double-masculine	3.66	.003	0.78	0.35, 1.20
Gender-neutral vs. Single-masculine	3.17	.015	0.67	0.25, 1.09
Gender-neutral vs. Double-masculine	2.84	.039	0.60	0.18, 1.01
Single-masculine vs. Double-masculine	0.35	.997	0.07	-0.34, 0.49
General Academic Competence (*df* = 221)				
Double-feminine vs. Single feminine	0.92	.890	0.19	-0.22, 0.61
Double-feminine vs. Gender-neutral	2.87	.036	0.60	0.19, 1.02
Double-feminine vs. Single masculine	3.58	.004	0.75	0.33, 1.17
Double-feminine vs. Double masculine	6.72	<.001	1.42	0.98, 1.85
Single-feminine vs. Gender-neutral	1.94	.298	0.40	-0.01, 0.81
Single-feminine vs. Single-masculine	2.66	.059	0.56	0.14, 0.97
Single-feminine vs. Double-masculine	5.66	<.001	1.19	0.76, 1.61
Gender-neutral vs. Single-masculine	0.74	.946	0.16	-0.26, 0.57
Gender-neutral vs. Double-masculine	3.90	.001	0.82	0.40, 1.23
Single-masculine vs. Double-masculine	3.12	.017	0.66	0.24, 1.07
Aptitude in Math and Science (*df* = 223)				
Double-feminine vs. Single feminine	2.39	.121	0.50	0.09, 0.91
Double-feminine vs. Gender-neutral	1.99	.273	0.41	0.00, 0.82
Double-feminine vs. Single masculine	6.33	<.001	1.33	0.90, 1.76
Double-feminine vs. Double masculine	10.82	<.001	2.26	1.80, 2.71
Single-feminine vs. Gender-neutral	0.43	.993	0.09	-0.32, 0.50
Single-feminine vs. Single-masculine	3.94	.001	0.83	0.41, 1.25
Single-feminine vs. Double-masculine	8.41	<.001	1.76	1.32, 2.20
Gender-neutral vs. Single-masculine	4.41	<.001	0.93	0.51, 1.35
Gender-neutral vs. Double-masculine	8.94	<.001	1.87	1.43, 2.32
Single-masculine vs. Double-masculine	4.45	<.001	0.93	0.51, 1.35
Career Suitability (*df* = 222)				
Double-feminine vs. Single feminine	4.21	<.001	0.89	0.47, 1.31
Double-feminine vs. Gender-neutral	5.62	<.001	1.18	0.75, 1.60
Double-feminine vs. Single masculine	11.72	<.001	2.47	2.00, 2.94
Double-feminine vs. Double masculine	13.60	<.001	2.85	2.36, 3.34
Single-feminine vs. Gender-neutral	1.36	.654	0.29	-0.13, 0.70
Single-feminine vs. Single-masculine	7.51	<.001	1.58	1.14, 2.02
Single-feminine vs. Double-masculine	9.37	<.001	1.96	1.51, 2.41
Gender-neutral vs. Single-masculine	6.24	<.001	1.31	0.88, 1.74
Gender-neutral vs. Double-masculine	8.11	<.001	1.69	1.25, 2.13
Single-masculine vs. Double-masculine	1.81	.368	0.38	-0.03, 0.79

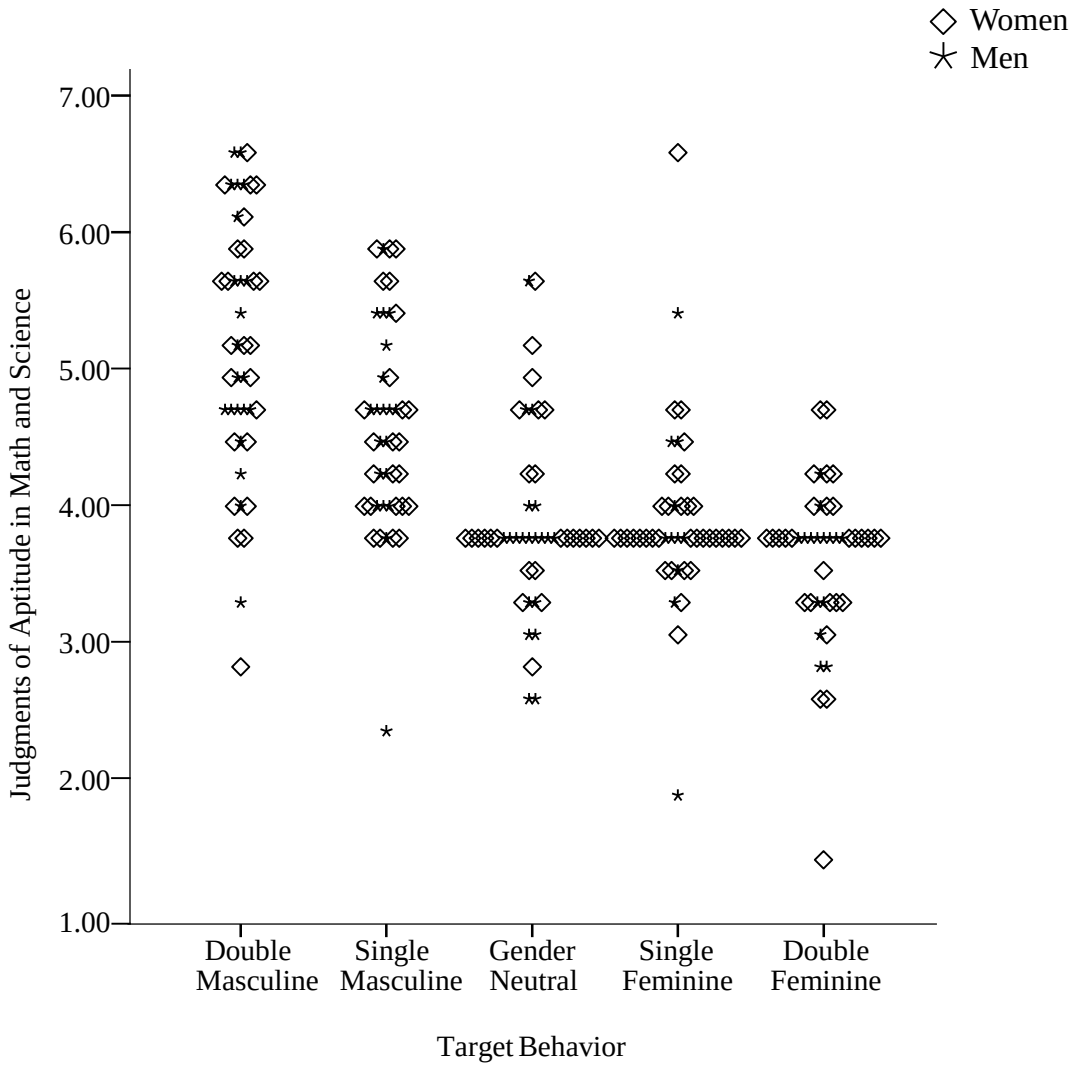
Note: Comparisons performed with Tukey's HSD.



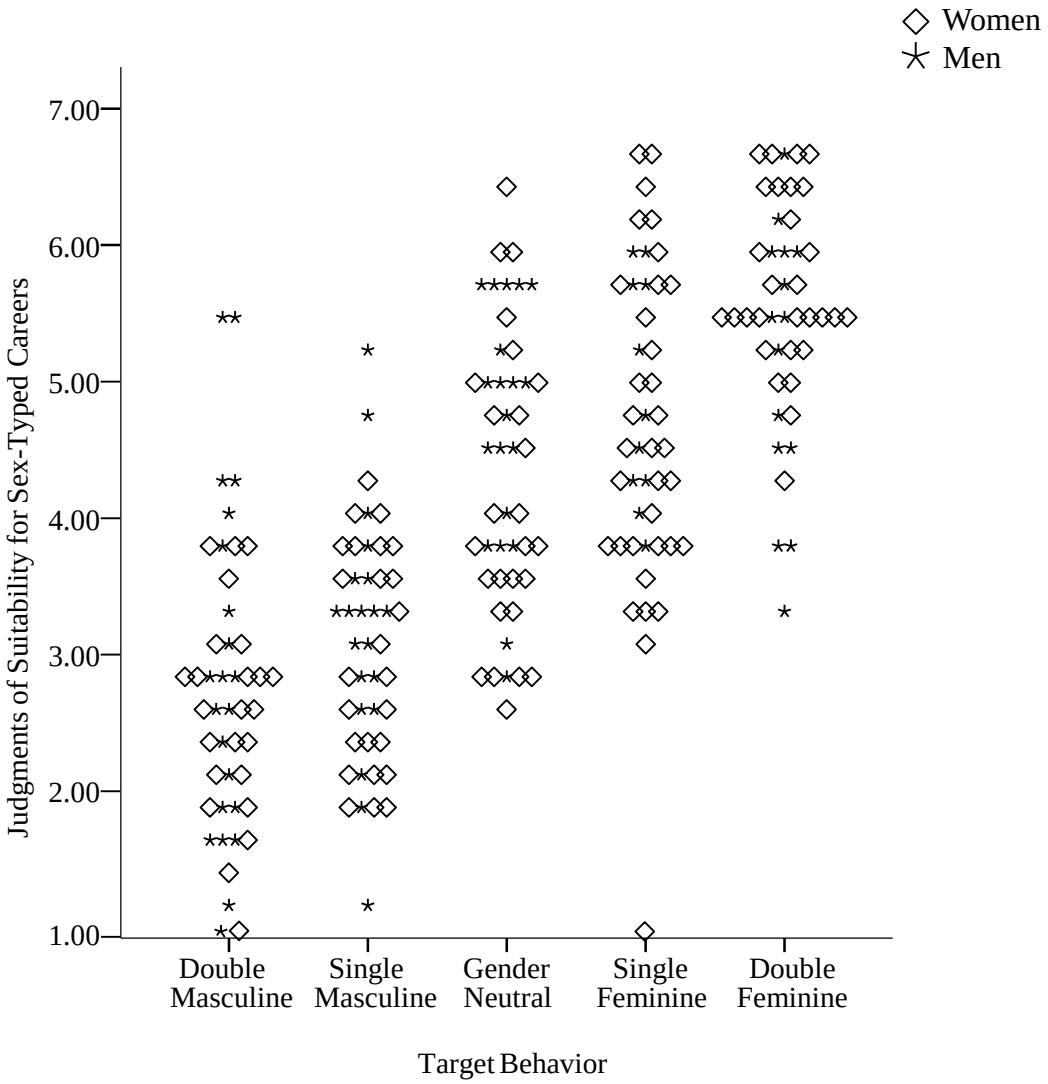
Supplemental Figure 1a. Experiment 4. Scatterplot of women's and men's judgments of the target's sex-typed traits.



Supplemental Figure 1b. Experiment 4. Scatterplot of women's and men's judgments of the target's academic competence.



Supplemental Figure 1c. Experiment 4. Scatterplot of women's and men's judgments of the target's aptitude in math and science.



Supplemental Figure 1d. Experiment 4. Scatterplot of women's and men's judgments of the target's suitability for sex-typed careers.