American Journal of Sociology

How Beauty Works. Theoretical Mechanisms and Two Empirical Applications on Students' Evaluations of Teaching

--Manuscript Draft--

Manuscript Number:	403930R1
Full Title:	How Beauty Works. Theoretical Mechanisms and Two Empirical Applications on Students' Evaluations of Teaching
Order of Authors:	Tobias Wolbring
	Patrick Riordan

How Beauty Works. Theoretical Mechanisms and Two Empirical Applications on Students' Evaluations of Teaching

Abstract

Plenty of studies show that the physical appearance of a person affects a variety of outcomes in everyday life. However, due to an incomplete theoretical explication and empirical problems in disentangling different beauty effects, it is unclear which mechanisms are at work. To clarify how beauty works we present explanations from evolutionary theory and expectation states theory and show where both perspectives differ and where interlinkage appears promising. Using students' evaluations of teaching we find observational and experimental evidence for the different causal pathways of physical attractiveness. First, independent raters strongly agree over the physical attractiveness of a person. Second, attractive instructors receive better student ratings. Third, students attend classes of attractive instructors more frequently – even after controlling for teaching quality. Fourth, we find no evidence that attractiveness effects become stronger if rater and ratee are of the opposite sex. Finally, the beauty premium turns into a penalty if an attractive instructor falls short of students' expectations.

Keywords: physical attractiveness, beauty premium, beauty penalty, status, sex, evolution, students' evaluations of teaching

Introduction

Many studies investigating how beauty might affect social outcomes have accumulated over the past years. This body of research brought forward theories, derived hypotheses, and empirically tested the ways that looks influence various areas of society such as the labor market (Hamermesh and Biddle 1994; Mobius and Rosenblat 2006), law and crime (Biddle and Hamermesh 1998; Mocan and Tekin 2010), trustworthiness and reciprocity (Andreoni and Petrie 2008; Mulford et al. 1998; Solnick and Schweitzer 1999), personal relationships and marriage (Elder 2008; Margolin and White 1987; Simpson, Gangestad, and Lermer 1990; Townsend and Levy 1990), public elections (Hamermesh 2006; Rosar, Beckers, and Klein 2008, 2012), school (Jackson, Hunter and Hodge 1995; Ritts, Patterson and Tubbs 1992), game shows (Belot, Bhaskar, and de Ven 2012), and even the use of space on sidewalks as a measure of dominance and power (Dabbs and Stokes 1975). The bottom line is that beauty is consistently relevant in shaping social relationships in all these studies, and hence it is an important dimension of social inequality. In light of cumulative *attractiveness treatment advantage* over the life course (DiPrete and Eirich 2006; Jæger 2011) the relevance of sociological inquiry into the "ugly stratification" of this “beautiful field” becomes abundantly clear.

The *beauty premium* regularly produces differences in at least the five following ways.

Beauty Consensus: Attractiveness ratings of different raters within a culture strongly correlate (Henss 1992; Patzer 1985, 2007). There are social aspects to current beauty standards and although attractiveness norms vary some with time and context, certain judgment criteria are remarkably stable over time and between cultures (Cunningham et al. 1995; Gangestad and Scheyd 2005; Grammer et al. 2003; Langlois et al. 2000; Langlois and Roggman 1990; Rhodes et al. 1998; Rhodes et al. 2002).

Attractiveness Attention Boost: Good-looking faces draw more attention than unattractive and averagely attractive ones and they are noticed faster and more frequently (Maner et al. 2007; Mulford et al. 1998). This is also true for infants who prefer attractive faces and spend more time looking at them (Langlois, Roggman, and Rieser-Danner 1990; Samuels/Ewy 1985).

Gender-specific attractiveness stereotypes: Attractive individuals trigger a number of general positive ascriptions. The stereotype activated may be summarized in the popular formula "what is beautiful is good" (Dion, Berscheid and Walster 1972: 285), since people regard attractive individuals as more sociable, kind, well mannered, honest, reliable, intelligent, creative, successful, and mentally and physically healthy (Eagly et al. 1991; Feingold 1992; Jackson et al. 1995; Langlois et al. 2000). Furthermore, individuals attribute the prevalent gender stereotypes to exceptionally good-looking females and males (Gillen 1981; Heilman and Okimoto 2007; Heilman and Saruwatari 1979).[1]

Attractiveness Glamour Effect: Attractive people have a higher chance of others attributing their mistakes to adverse circumstances. This leads to a decreased likelihood of being personally blamed and sanctioned by relevant others (Bassili 1981; Dion, Berscheid, and Walster 1972). Thus, even blatant misconduct of attractive

[1] These gender-specific attractiveness stereotypes can have negative consequences known as the beauty-is-beastly effect (Heilman and Saruwatari 1979). In a context demanding masculine characteristics, such as management, typically feminine attributes will usually lead to inferior treatment resulting in a negative beauty bias. Good-looking men, on the other hand, are expected to be treated worse than their less beautiful male colleagues in areas, such as jobs in kindergartens where male competitiveness is less advantageous than female empathy.

individuals does not necessarily impair their "aura" of being beautiful and, therefore, good.

Beauty Penalty: Physical attractiveness is not advantageous under all circumstances. Some studies indicate that misbehavior by persons with a stunning appearance can cause stronger sanctions (Andreoni and Petri 2003; Wilson and Eckel 2006). Research by Sigall and Ostrove (1975) suggests that such reactions are more likely if attractiveness is a relevant characteristic for the task under study (see also Webster and Driskell 1983: 142).

Explaining this bundle of findings is the major theoretical task in research on the effects of physical attractiveness in social interactions. Although most existing research refers to at least some of these empirical regularities, the social mechanisms underlying these phenomena are, however, rarely explicated within a coherent theoretical framework. In this paper we draw on evolutionary theory and expectation states theory to address competing and complementary explanations. We propose mechanisms that could bring about the beauty premium and penalty, and empirically test them in one selected area: *students' evaluations of teaching* (SET) (Hamermesh and Parker 2005; Klein and Rosar 2006; Süssmuth 2006; Wolbring 2010a). The main finding of these previous studies is that better-looking teachers receive better evaluations than their less beautiful colleagues, all else equal.

Upon closer inspection, the advantages of studying SET to research effects of physical attractiveness in general become clear. The central problem of empirical studies on the effects of physical attractiveness is causality, since the beauty effect might be confounded with a performance effect. For example, in labor market studies it is seemingly impossible to control for all aspects of productivity in order to isolate the causal effect of physical appearance. In comparison, studies on SET offer three fundamental advantages: First, using SET as a quality indicator and controlling for other potentially confounding variables (e.g., characteristics of

course, student, and teacher) we can discriminate between the effects of teaching quality and the instructors' appearance. Although SET obviously do not yield perfectly objective measures of teaching quality, students' subjective ratings capture a lot about teaching quality (for broad overviews see Marsh 2007; Spooren, Brockx, and Mortemans 2013). Second, the familiarity of students with teaching, test situations, and SET facilitates the use of laboratory experiments, which help to further disentangle effects of productivity and discrimination with a high degree of internal validity. By systematically varying physical appearance while holding teaching performance constant we can validate the findings from the observational study. Third, the laboratory setting allows us to test the proposition of the beauty penalty, which is quite difficult with observational data. Due to ethical restrictions, students cannot be issued tests of varying difficulty in actual teaching situations thus we have to rely on fairly vague indicators of course difficulty in our analysis of observational SET data.

We combine the strengths of both approaches: Study 1 draws on panel data of actual SET combined with exogenous measurements of the instructors' physical attractiveness. This study specifically allows us to assess the effect of the instructor's appearance on course ratings while controlling unobserved heterogeneity at the student level thanks to repeated observations of the same students across different classes. Moreover, holding course quality constant we ask whether students attend classes held by attractive teachers more frequently. Study 2 compliments these observational analyses with a series of laboratory experiments, in which SET for a fictitious lecture including a test were gathered. Treatments are the instructor's physical attractiveness and sex, the difficulty of the test, and the sequential arrangement of test and SET. This setup allows us to test the hypothesis on the possibility that the beauty premium turns into a penalty under certain circumstances.

Theoretical background

The following section consists of theoretical arguments from evolutionary theory and expectation states theory. First, we present the main arguments from evolutionary theory, especially from sexual selection theory proposing that beauty signals fitness and hence corresponding preferences yielded evolutionary advantages. Although applying evolutionary arguments to modern and not explicitly reproductive contexts of SET appears far-fetched, the evolution of this selective strategy might have persisted and spread to non-mating tasks such as teaching. Second, we discuss expectation states theory which treats beauty as a diffuse status characteristic highlighting the self-enforcing dynamics of social interactions. At the end of these theoretical considerations, we discuss the implications of both theories, elaborate on the basic differences and respective black boxes of the two frameworks, and expound where theoretical interlinkage appears promising.

Evolutionary basis of preferences toward beauty

The relevance of evolutionary theory and genetics for sociological research has been discussed extensively (Freese 2000; Freese and Shostak 2009; Guo 2006; Hopcroft 2009; Kanazawa 2001; Lenski 2005; Lopreato and Crippen 2002; Nielsen 1994; Richerson and Boyd 2004; Udry 1995). As evolutionary psychology argues (Barkow, Cosmides, and Tooby 1992; Buss 1999, 2005; Cosmides and Tooby 1989; Pinker 2003), the main reason is that humans and their brains are products of a long process of evolution which consists of two primary mechanisms: natural selection and sexual selection (see Mayr 2001 for a highly accessible introduction into evolutionary theory).

Natural selection focuses on the survival of individuals based on their fitness relative to the surrounding environment. Yet being perfectly adapted to survival is useless, if no offspring are produced. Sexual selection, therefore, is seen as crucial for success or failure of strategies

in the long term; individual orientation towards fitness maximization is not the result of a conscious decision, but of evolutionary processes (Schulmann 1978; see also Lopreato and Crippen 2002:120).

Because an individual's genes are recombined with a mate's genes to produce the genotype of the offspring, it seems imperative that the chosen mate is as fit as possible. Since explicit information on the genetic fitness of a potential mate is not readily available individuals' evolved strategies must rely on those fitness signals at hand, mainly the phenotype of a potential mate. This is what Freese (2008) termed the "phenotypic bottleneck": individuals can only rely on the looks and observable actions of others. Since physical attractiveness is correlated with youth, fertility, and health (Buss 1999; Feingold 1992; Grammer et al. 2003; Møller and Swaddle 1997; Rhodes et al. 2001), and in some studies also with intelligence (Jackson et al. 1995; Kanazawa 2011), evolutionary theorists argue that beauty is a valid signal of fitness (see Gangestad and Scheyd 2005). Although the indicator may be noisy, the signal is hard to manipulate and easily observable at next to no cost (e.g., Miller and Todd 1998). Hence, from a mating perspective it is obvious why physical attractiveness matters. For the same reason – having beautiful children results in a higher chance of reproducing one's own genes – the theory predicts that parents will invest more resources in attractive than in unattractive offspring bringing about differences in ability and leading their children to be more successful in non-reproductive domains.

Hence, choosing attractive mates is argued to be an evolutionary rational strategy, which successfully replicates and spreads over time: Those individuals able to detect physical attractiveness, and thus either high-quality mates or serious rivals quickly were able to survive and reproduce more successfully. Wilson und Eckel (2006: 2000), therefore, argue that "attentiveness to attractiveness may be embedded as part of our cognitive apparatus".

Upon closer inspection it becomes obvious that, from an evolutionary perspective, attractiveness research must take sex into consideration. According to parental investment theory (Trivers 1972), physical attractiveness has different consequences for males and females since they contribute to reproduction in quite dissimilar ways. Evidently, even in modern cultures females bear considerably higher costs than males: not being able to mate while enduring the strain and risks of pregnancy and birth, as well as the subsequent nursing period, which requires a higher intake of calories and special attention to the child. For males, on the other hand only a few minutes of copulation is technically required in order to pass on their genes.

This disparity in the investment required for offspring is hypothesized to lead to very different strategies by sex when competing for mates (*intrasexual competition*) and when trying to impress the most desirable members of the opposite sex (*intersexual selection*). Since females incur higher costs, they are suspected to be a lot more "choosy" when it comes to mate selection. Males, in contrast, will be in much more fierce competition with each other for access to desirable females. Although females also compete for the fittest males and males may be cautiously selective, evolutionists characterize human mating as mostly dominated by male competition and female choice (Buss 1999; Miller 2001).[2]

[2] Concerning intersexual selection, another differentiation is necessary: the potential length of the mating relationship. Human mating can span various periods of time ranging from only a few minutes of intercourse (*short-term mating*) up to many years of joint child rearing (*long-term mating*). The relative importance of evaluation criteria for potential mates might shift with the length of the mating relationship (Buss 1989, 1999; Miller 2001). In long-term mate selection, individuals (of both sexes) would primarily seek signals for characteristics which help fitness maximization in the long run. Females would, thus, look for males with resources and the willingness to invest them in joint offspring. Males would place higher importance on fidelity, a trait that reduces the risk of investing

From these evolutionary arguments about the most desirable mate qualities, attractiveness is one relevant trait for female selection (females choosing males), but it is a lot more important in male selection (males choosing females). Further, attractiveness might be relevant in intrasexual competition, but since the driving force behind the evolution of signals for genetic quality is mate selection, attractiveness will be more important between sexes. From a mating perspective this *intersexual attraction effect* is thus one centerpiece of the evolutionary framework.

Beauty as status

Thus far our argument mainly concerned human reproduction and genetic survival. Although one could argue that mating and other social outcomes such as students' evaluations of teaching are unrelated domains, numerous aforementioned studies have shown physical attractiveness is also associated with non-reproductive outcomes in diverse fields such as the labor market, law, and public elections. It is possible to transfer the evolutionary argument to these contexts: As is well known, cognitive resources of humans are limited and information on interaction partners is usually incomplete in all domains of social life (Buss 1999; Gigerenzer 2008; Simon 1957). Nonetheless, humans must constantly make decisions and, therefore, in order to remain capable of action they use information shortcuts and heuristics. One such decision shortcut might be to transfer strategies which were successful in one context (e.g. mating) to other settings.

---

resources in another man's offspring. In the short-term case, both sexes are argued to focus on maximizing the genetic quality of their offspring. As discussed above, attractiveness is a general fitness signal in mating, but especially in short-term mating. Because of their greater investment requirements, females are posited to apply a long-term strategy in most (but not all) situations while men would be more likely to follow a short-term approach (Buss 1989; Symons 1979; Trivers 1972).

Another heuristic which might be a product of evolution is to rely on social status as a signal for influence and desirable outcomes. Since dominance hierarchies are a common feature of most animal and human groups (see Chase and Lindquist 2009), from an evolutionary perspective individuals take into account the relative position of their interaction partners (see also Wilson and Eckel 2006: 201). Although physical attractiveness need not be strongly correlated with classical measures of status, namely "exchangeable resources" such as education, wealth, and income ("consummatory value goal objects"),[3] it indicates access to other desirable resources like friendship, attention, and opinion leadership ("status value goal objects" ; see Hysom 2009; Webster and Hysom 1998: 361). The reason for this widespread belief is that people recognize where advantages and disadvantages in social interactions flow to in everyday life. Thus physical appearance can be viewed as a salient characteristic in social interactions: Seeing beautiful or ugly people (of the same or the different sex) activates social categorization and stereotypes which influence social interactions.

These stereotypes can have an evolutionary basis, whereby mankind's "experiences" regarding genetic survival are hard-wired in the human brain and are (rightly or wrongly) used in non-reproductive domains of life. Moreover, since humans probably regard phenotypic looks as a valid indicator for social status and success, they might follow and imitate beautiful people and treat them better.

Besides these ultimate causes, proximate causes such as cultural processes and social interactions shape and change stereotypes of attractive and unattractive people (Rosar 2009;

[3] McClintock (2014) reports for small but significant correlations of physical attractiveness with years of completed education, expected college graduation status, projected future occupational status, and projected future income in the US. In addition, Mueller and Mazur (1996) find that facial dominance is a predictor of military rank in the late career of West Point Cadets.

Synnott 1989, 1990; Webster and Driskell 1983). For example, people learn from books, films, songs, and the behavior of others that physical attractiveness and beautiful traits are positively associated. Thereby, social stereotypes can, but need not necessarily be related to evolutionary influences. Due to differential expectations and treatment small initial differences (even if they are random) in traits, abilities, and/or available resources between attractive and unattractive people are sufficient to bring about huge distributional inequalities over time.

Expectation states theory (Berger et al. 1977; Ridgeway 1991; Ridgeway et al. 2009; Webster and Hysom 1998) offers a general framework for such considerations on the self-enforcing and self-fulfilling nature of hierarchical orderings in social encounters. The main assumption of this theoretical framework is that nominal characteristics, such as gender, race, and occupational position, do not signal genetic fitness, but serve as diffuse status characteristics for prestige and power. More precisely, a *diffuse status characteristic* is defined as (1) being socially significant in a given culture and (2) encompassing nominal states, which discriminate between members of the population and which go hand-in-hand with (3) general and differentially valued expectations about traits, capacities, and behaviors (for similar, but not identical definitions see Berger, Cohen, and Zelditch 1972; Berger and Fişek 2006; Webster and Driskell 1983). Since physical attractiveness meets these criteria for a diffuse status characteristic, expectation states theory applies to the case of beauty. Webster and Driskell (1983) explicate that beauty is an easily observable and salient category in everyday interactions allowing clear differentiation between members of the population, and eliciting general expectation states for attractive and unattractive individuals that are relevant for *most* tasks.

Expectation states theory is unfortunately less clear about how attractiveness expectation states combine with stereotypes for additional diffuse status characteristics. Counter to the

intersexual attraction hypothesis, Webster and Driskell (1983: 154) point out and provide empirical evidence that beauty and gender effects combine additively but *do not interact* (see also the review of empirical evidence by Morrow 1990). Meanwhile, proponents of status construction theory emphasize that additional status categorizations, such as physical attractiveness, are "cognitively nested within the fundamental understanding of the person as male or female" (Ridgeway and Smith-Lovin 1999: 193). In addition, some argue that beauty is more salient and influential for females than males in modern Western cultures – especially (but not only) if males evaluate females in social encounters (Wolf 1991). Hence, while expectation states theory and evolutionary theory clearly differ in their conceptualizations of sex and gender and the respective theoretical underpinning of beauty effects, both frameworks do not necessarily disagree in their predictions of an intersexual attraction/cross-gender effect.

Irrespective of whether attribution is gender-specific or not, expectation states theory proposes that actors presume attractiveness stereotypes as given and informative: "The burden of proof is placed upon demonstrating that status is *not* relevant to ability, instead of the other way around" (Webster and Driskell 1983: 146). In other words, the theoretical assumption is that, if not clearly proved otherwise, status is always relevant in the situation at hand. This can even blind people into attributing failures of attractive individuals to external sources (see attractiveness glamour effect). Perceived status not only influences performance expectations, but also biases cognition and performance-evaluations.

Expectation states theory further posits that status organizes and channels social interactions. People act upon their general expectation states resulting in differential treatments of others. Since "individuals appear to form performance expectations for actors with given characteristics on the basis of their past experience with actors who have those characteristics" (Ridgeway 1991: 375), social encounters with stereotype inconsistent attractive and unattractive others do not necessarily lead to the reduction of stereotypes.

Moreover, interactional processes often cause the maintenance and spread of socially constructed status beliefs because individuals preform attributions, and therefore manufacture their own fulfillment of stereotypes. This can be viewed as a primary micro process for the accumulation of social inequality over the life course.

Taken together, evolutionary and expectation states theories predict a *beauty bias*, i.e. differential treatment of persons with different physical attractiveness independent of their skills or productivity. This is the case when people erroneously regard physical attractiveness as a valid signal or simply prefer good-looking people irrespective of other individual characteristics. As shown above, this beauty bias usually takes the form of a *beauty premium*.

Physical attractiveness and other diffuse status characteristics do not have positive consequences under all circumstances. Instead, individuals may be punished for their good looks in certain situations (*beauty penalty*). As mentioned expectation states theory assumes that beauty always matters and that the burden of proof lies on demonstrating that a given status characteristic is *not* a valid signal for traits of interest. Thus, due to positive reinforcement and attractiveness glamour effects expectations usually are self-confirmatory and stabilizing over time. However, the framework leaves out the question of what happens in a situation where there is overwhelming non-negligible proof that a subject does not meet the expected qualities.

Insights from social psychology suggest that when faced with clearly disconfirming actions people try to restore consistency (e.g., Festinger 1957). Thus, if failures are too severe to either simply overlook or attribute to external adverse conditions, individuals will adjust their originally positive image of an attractive person downward. Moreover, blatant frustrations of expectations will engender disappointment and anger as mechanisms intended to prevent similar experiences in future encounters. Negative reactions are especially likely if the

onlookers are directly affected by negative consequences of a counter-stereotypical behavior. A well-known behavioral principle to punish violations of expectations in such cases is direct reciprocity. Various classical scholars have pointed to this regularity in social life (Blau 1964; Gouldner 1960; Malinowski 1922; Mauss [1923] 1990; Simmel 1900, 1908) and numerous empirical studies in anthropology, biology, economics, psychology, and sociology (see reviews by Adloff and Mau 2006; Bowles and Gintis 2011; Kolm and Ythier 2006) have shown that humans tend to treat others similar to the way these others have treated them (*quid pro quo*).[4] This is especially true, if the observed behavior is interpreted as an act of intentional friendliness or hostility.

Thereby, type and strength of reactions to *blatant misconduct* (i.e. the failure to fulfill expectations) will depend on the lack of fit between expectations and behavior (see also Darley and Fazio 1980; Zebrowitz 1997): the more performance deviates from social stereotypes, the higher is the contrast between expectations and observed behaviors. Consequently this evokes greater frustration and subsequent increased desire for stronger sanctioning of the target person. Hence, for the case of beauty we expect good-looking individuals to be punished more severely for clear violations of behavioral expectations since appealing visual cues raise the expectations in the behavior of such individuals, attractive others are anticipated to perform better, commit fewer errors, and abide by norms in a

[4] Reciprocity can be a rational strategy for individuals and groups in terms of survival and reproduction and, thus, fits squarely within an evolutionary framework. It can be individually advantageous to behave in such a way since it encourages cooperation while simultaneously deterring defectors in a rather efficient way (Axelrod 1984, 1997). As well, there are different rationales for the use of this principle in groups. For example, Trivers (1971) has argued that groups whose members helped each other out and expected the favor to be returned at a later stage (reciprocal altruism) were more successful, allowing their members to survive and reproduce.

superior way. Thus, obvious failures cause a higher degree of frustration on behalf of onlookers who desire restoration of congruency with the expected state. Experimental findings by Andreoni and Petri (2003) and Wilson and Eckel (2006) corroborate this hypothesis: In public goods and trust games attractive individuals are initially treated more favorably. However, if information on untrustworthy and uncooperative behaviors of attractive individuals is provided, the beauty premium turns into a penalty. Hence, these findings suggest that disappointments by highly esteemed (attractive) persons actually cause stronger sanctions than misbehavior conducted by less important individuals.

Comparison of the theoretical perspectives

Evolutionary theory and expectation states theory have a lot to contribute to the analysis of the effects of beauty. While evolutionary theory is mainly concerned with the replication of strategies and fitness enhancing traits and behaviors, expectation states theory centers on the concrete workings of social interaction regarding the reproduction of status. Given the stark differences between these theories it is remarkable that both lead to very similar expectations concerning how beauty works. Evolutionary theory and expectation states theory are not mutually exclusive which is why we do not attempt to test these theories against one another. Instead we expect to better understand beauty by studying how their different foci interact, overlap, and contrast. We propose that each can shed light inside the theoretical black box of the other.

The “stopping rule” (Miller 1987) of expectation states theory concerns the origin of beauty as a status characteristic. Expectation states theory leaves this aspect blank while making the important point that the origin of initial differences is negligible in the face of the strong reinforcement mechanisms building on them. While this is theoretically parsimonious, the question of origin is important when searching for an explanation of the effects of beauty.

Why is physical attractiveness a salient characteristic in social interactions and why do we attribute positive characteristics to good-looking people? Social psychological theories often locate origins of status characteristics in culture and social constructionism. This offers proximate causes but leaves the deeper question of why specific stereotypes are the way they are unanswered. Evolutionary theory adds a rationale for the ultimate causes of a preference for beauty and helps us to understand why certain characteristics of the body are perceived as attractive and others are not.

On the other hand, evolutionary theory does not elaborate on the self-enforcing dynamics of social interactions. Thus, "a main role for genetics is as a placeholder for ignorance of more proximate influences of psychological and other embodied variation" (Freese 2008: S1). Expectation states theory can open this genetic black box by offering a better understanding for the social processes at hand. Thus, as other authors have previously emphasized in this journal evolutionary theory does not undermine existing sociological perspectives because "genetic expression can only reveal itself through social structural change" (Bearman 2008: SV) and "dynamics of social interaction and organization affect the influence of genes on outcomes" (Freese 2008: S25).

Concerning the role of sex/gender both theories have a soft spot. From an intersexual attraction perspective, it seems clear that the evolutionary framework implies stronger beauty effects between sexes, but upon closer inspection in regard of intrasexual competition this implication loses some of its discriminatory power: Evolutionary theory does not rule out beauty effects within one sex, since physical attributes are valuable signals of fitness in potential competitors and allies. Expectation states theory faces a similar problem, since it does not make clear how gender interacts with other status characteristics.

Having clarified similarities and differences between evolutionary theory and expectation states theory and their corresponding mechanisms we want to test some of their empirical implications for a selected area of research, the evaluation of teaching. Before presenting the research strategy and results we specify our theoretical considerations with regard to this field and give a concise overview of attractiveness research on SET.

Attractiveness and Students' Evaluations of Teaching

Generally, empirically exploring the causal effects of physical attractiveness and testing the corresponding hypotheses has proven fairly difficult. The main reason for this is that physically attractive and unattractive humans often do not only differ in their appearance but also in other characteristics relevant for the outcome of interest. For example, if physically more attractive employees receive higher incomes this could be the result of (a) an attractiveness treatment advantage, (b) spurious correlation due to a positive association between looks and factors influencing productivity such as self-confidence, intelligence, and ability, or (c) reverse causality: higher incomes allow purchasing of a certain degree of beauty, e.g. through nicer clothes, makeup, healthier diets, gym memberships, and personal trainers (see Hamermesh, Xing, and Zhang 2002).

Besides the practical relevance of research on SET for the design of higher education systems the "beauty" of this field lies in the possibility of generating measures of physical attractiveness of instructors, the quality of their work, and students' perceptions of the actual teaching quality of instructors. Because of this, SET offer a great opportunity to disentangle the effects of discrimination and productivity more clearly.

Most empirical studies concentrated on students' global ratings of course quality or teaching performance such as "All in all, how do you rate the overall quality of this course?". These measures are commonly taken to be indicators of teaching quality in students' perceptions and, therefore, should reflect the attractiveness treatment advantage if it exists. However, an

association between physical attractiveness and teaching quality can also point to differences in productivity and, hence, in actual teaching quality. More attractive instructors might receive better ratings because of their higher presentation skills or because students pay closer attention. This would lead to a higher learning success in courses taught by good-looking instructors, all else equal. Controlling for such influences by means of an experimental design or in statistical models is the most popular way to circumvent those difficulties. Additionally it appears important to test the theoretical mechanisms at hand to gain more confidence in the causal nature of one's results.

Since students might not only use appearance as a diagnostic tool to infer teaching quality but might also simply prefer to interact with attractive others they probably attend classes of attractive instructors more frequently even if they are not – or are not perceived as – better teachers. Therefore and in contrast to previous studies, we regard students' attendance behavior as an additional important outcome variable.

Moreover, identifying an intersexual attraction effect and a beauty penalty would further increase our confidence in a causal effect of physical attractiveness on SET. Giving more weight to looks of instructors simply because they are female and sanctioning them more strongly for failures simply because they are physically attractive cannot be justified solely by differences in productivity. Thus, the proposed mechanisms help to separate between the competing explanations. In the following we want to test five hypotheses in order to shed more light on the processes at hand:

H1: There is a high degree of interindividual agreement on the physical attractiveness of instructors among students. (*Beauty Consensus*)

H2: Students skip classes less when instructors are physically more attractive even if course quality is held constant. (*Attractiveness Attention Boost*)

H3: Physically attractive instructors receive more favorable evaluations of teaching. (*Attractiveness Treatment Advantage*)

H4: The effect of physical attractiveness on course ratings and absenteeism is stronger, if the sexes of rater and ratee differ. The effect is more pronounced if a male student rates a female instructor. (*Intersexual Attraction Effect*)

H5: Greater physical attractiveness of an instructor leads to worse ratings in the case of her failing to meet the expectations of students. (*Beauty Penalty*)

State of Research. In previous studies on SET only hypothesis 3 has been tested. Furthermore, most studies (except Wolbring 2010a) explore an interaction effect between the instructors' sex and their physical attractiveness, but not the possibility of an interaction with the sex of rater and ratee. In general, researchers relied on three different empirical strategies to identify effects of physical attractiveness and their interaction with sex.

(a) In laboratory experiments Buck and Tiene (1989) and Ambady and Rosenthal (1993) simulated a teaching situation using audio or video tapes. They found no or only weak overall effects of attractiveness, whereas a similar study by Goebel and Cashen (1979) demonstrated that beauty has a clear influence on students' ratings of teaching competence. Moreover, Buck and Tiene (1989) report an effect that is stronger for female instructors, contrary to results by Goebel and Cashen (1979) that do not point in this direction.

(b) Others recently used data from online platforms such as www.ratemyprofessors.com and found very strong effects (Felton et al. 2004, 2008; Riniolo et al. 2006; Rosar and Klein 2009). Attractiveness of an instructor can improve her ratings by up to 2.5 out of 5 grade points. Moreover, two of these studies test the possibility of an interaction with the sex of the instructor and find mixed results. Riniolo et al. (2006) find no sex-

specific differences, whereas Klein and Rosar (2009) support the hypothesis of a stronger effect for female instructors. However, these studies have two major drawbacks (Wolbring 2010b): First, with the exception of Klein and Rosar (2009), they are based on potentially endogenous measures of physical attractiveness because students evaluated appearance and quality of teaching simultaneously. Second, all studies rely on a self-selected sample whereby the selection process can be driven by the variables of interest. Both methodological difficulties presumably lead to an overestimation of the actual effect.

(c) The most reliable evidence with regard to external validity stems from research using actual SET data and exogenous ratings of physical attractiveness (Hamermesh and Parker 2005; Klein and Rosar 2006; Süssmuth 2006; Wolbring 2010a). All of these studies find significant effects that are markedly weaker than those reported in studies of type (b) but are nonetheless non-negligible and stronger than in studies of type (a). With studies finding no clear sex-specific differences in the effects of looks (Klein and Rosar 2006; Wolbring 2010a) and others showing stronger effects for male instructors (Hamermesh and Parker 2005), the evidence on interaction with the sex of the instructor is again inconclusive. Moreover, the results by Wolbring (2010a) do not corroborate the intersexual attraction hypothesis.

In addition to our theoretical explanation for the effects of physical attractiveness this paper adds to the literature on SET and illustrates reasons why research on beauty biases can benefit from work focusing on SET. However, before presenting the empirical evidence we want to clarify our identification strategy for the causal effects of beauty.

DAGs for Attractiveness Effects. Following the suggestion by Judea Pearl (2009) we depict our theoretical arguments in directed acyclic graphs (DAGs) to shed light on the possibilities and limitations in statistically isolating the causal effects of beauty (for easily accessible

introductions see Elwert 2013; Morgan and Winship 2007). Figure 1 illustrates the major difficulty of our empirical endeavor: Since ability and looks of instructors could be correlated an unconditional estimate of the effect of physical attractiveness on SET is potentially confounded by differences in productivity. Therefore, the most obvious identification strategy is to directly control for instructors' abilities. However, using observational data this is nearly impossible since instructors' abilities are difficult to capture – usually they are measured via some form of SET which is ironically the outcome variable of interest. This is why we designed an experiment in which we can hold constant teaching quality while varying instructors' looks.

[Figure 1. Directed Acyclic Graph for Physical Attractiveness, SET, and Attendance]

A second strategy to isolate causal mechanisms is to focus on an additional outcome variable. As the DAG illustrates, in the situation at hand we can use students' class attendance as an additional outcome. We assume that attendance is influenced by students' perceptions of teaching quality (measured by SET), their interest and motivation, and instructors' physical attractiveness. The latter hypothesis is related to the attractiveness attention boost: ceteris paribus good-looking people receive more attention from their social environment. Once enrolled, students will thus skip courses taught by attractive instructors less often due to a more positive experience they make when attending an attractive teacher's class. The advantage of this approach is that one can now indirectly control for (unobserved) differences in ability by adjusting for SET and rule out one important pathway in the DAG. After conditioning on this backdoor path only one undesirable path remains: if instructors' abilities are correlated with their physical attractiveness and if abilities influence students' interest and motivation the estimate could still be confounded even if we control for subjective quality perceptions. Nonetheless finding an effect of physical attractiveness on attendance after conditioning on SET should strengthen the confidence in the causal nature of the effect.

The same is true for the third indirect identification strategy: Testing for two theoretically derived interaction effects. As explicated above from an evolutionary perspective one would expect that the effect of looks is stronger if rater and ratee are of the opposite sex. Although the estimates for the two cases (same sex; opposite sex) could also be confounded by differences in productivity it is extremely unlikely that there is no causal effect of beauty if the interaction is in accordance with the theoretical proposition. This also holds for the beauty penalty: if physically attractive instructors who fail the expectations of their students are rated worse than instructors who violated the same norm but are less attractive, the likelihood of a causal effect of beauty increases.

Using two different data sources we try to isolate the causal mechanisms of beauty from influences of productivity. In order to achieve a high degree of external validity we first analyze observational data from SET. Due to the panel structure of these data and the availability of numerous student and course characteristics we can control for more potential confounders than previous studies and can test for intersexual attraction effects (identification strategy 3). Furthermore, information on students' attendance behavior is available to empirically examine whether attractiveness boosts attention even if subjective teaching quality is held constant (identification strategy 2). However, common drawbacks remain (measuring physical attractiveness and teaching ability) and hypotheses on a beauty penalty can only be tested on the basis of fairly vague indicators of course difficulty and with low statistical power. This is why we designed study 2, an experiment complementing study 1 which allows us to directly control for productivity (identification strategy 1) and which offers insights on students' reactions to violations of behavioral expectations by their instructors (identification strategy 3).

Study 1: Attractiveness in Real Students' Evaluations of Teaching

Study 1 draws on extensive data from SET collected regularly at the Faculty of Social Sciences of the University of Munich. These data include information for the course (especially the students' overall ratings of the course), the person teaching (including an exogenous measure of physical attractiveness), and characteristics of those evaluating (sex, attendance). This allows us to explore, among other things, the effects of the instructors' physical attractiveness on students' ratings and class attendance.

Data and Research Strategy

In total approximately 18,000 observations for over 680 courses with at least 10 participants are available for the years 2008 to 2010 (covering 4 terms). Besides specific items on a variety of teaching quality dimensions and instructor behaviors the Munich SET questionnaire contains a global rating question on course quality ("On a continuous grading scale from excellent [5.0] to insufficient [1.0]: How do you rate this course?"). Moreover, for each course the number of classes missed is also assessed in the SET form ("Up to now, how many classes of this course have you missed?").

In order to measure the physical attractiveness of the instructors we collected portrait photos available on personal homepages and their department websites. One has to acknowledge that the attractiveness of a face is only one aspect of the broader theoretical construct 'physical attractiveness'. The latter also includes further dimensions such as height, weight, body shape, and gesture. Nonetheless for our purpose it is sufficient to focus on the face because of three findings. First, facial attractiveness is the most important dimension of overall physical attractiveness and is strongly correlated with the latter (Currie and Little 2009; Rhodes and Zebrowitz 2002; Snyder, Berscheid and Matwychuk 1985; Zebrowitz 1997). Second, the effects of face and body do not interact; both factors independently affect overall attractiveness and hence can be analytically isolated (Peters, Rhodes, and Simmons 2007).

Third, although different dimensions of the face (symmetry, plain skin, averageness) influence our perception of beauty (Gangestad and Scheyd 2005; Grammer et al. 2003; Rhodes 2006), the distinction between attractive and unattractive faces can be empirically established quite easily: Asking a group of raters for a judgment about portrait photos and using the average rating is a common way to measure physical attractiveness which is well-known as the "truth of consensus method" (e.g., Henss 1992; Patzer 1985, 2007).

Excluding courses with two or more instructors we were able to retrieve 177 photos of the 277 instructors in our data (63.9 %) covering 79.6 % of all observations in our data set.[5] 11 male and 9 female students from the University of Bern and 31 male and 34 female students from the University of Leipzig, who did not know the instructors, rated the photos on an 11-point-attractiveness-scale (very unattractive [0] – very attractive [10]). For analyses we use

[5] As can be easily seen, instructors for which photos were available teach more or bigger classes. Therefore, instructors with tenure are overrepresented in our data. Thus, one might be worried that this and other selection effects bias the estimates for the effects of physical attractiveness. For example, the fact of publishing a photo online might on average be associated with personality traits that also positively influence teaching (e.g., extroversion, psycho-social adaptivity). However, we find no systematic differences in SET ratings between instructors with and without internet photos. Moreover, measurement error might be an issue, since pictures varied in quality, age, and perspective. Instructors have probably chosen photos for their website which emphasize their attractive bodily characteristics and mask the unattractive ones. This could lead to a decrease in the variation of attractiveness measurements and a loss of precision for regression-based effect estimates. However, we empirically find significant differences in students' ratings of instructors' attractiveness and also significant attractiveness effects. Although this does not rule out the possibility of measurement error and biased estimates, it at least shows that sufficient variation in physical attractiveness remains even if measurement error is present.

each instructor's average rating as an exogenous measure for her physical attractiveness. Our analyses show that raters highly agree in their ratings (Cronbach's α = .95) supporting the findings of previous studies and hypothesis 1.[6] Since we are not only interested in the main effect of this variable but also in interaction effects we centered the variable at the mean.

Perceived course quality and absenteeism are the dependent variables in our regression models. For both outcomes we estimate simple linear regressions. Since the number of classes missed is a typical case of count data with overdispersion (Cameron and Trivedi 1998; Long 1997), we ran negative binomial and zero-inflated negative binomial models as sensitivity analyses. Since the results are very similar and the predicted values from both count data models are highly correlated with those from the simple linear regression model (r = .977 and r = .942; p < .001), we only report findings from the latter.

To take into account the fact that students usually complete more than one SET questionnaire each semester and to control for unobserved interindividual heterogeneity, which often biases results from cross-sectional analyses, we include student fixed effects in all our analyses (Allison 2009; Angrist and Pischke 2009; Wooldridge 2002).[7] We can do this because the

[6] Furthermore, we conclude from these results that our measurements are reliable. In order to give students the chance to get used to the rating task and to further validate our measures we first showed all students 12 pictures of male and female individuals strongly varying in their physical attractiveness. We took these photos from www.spin.de – a website where registered users can rate portrait photos. Correlating our ratings with evaluations from this website shows very high agreement (r = 0,948 [p < .001] between spin.de-raters and students from Bern; r = 0,954 [p < .001] between spin.de-raters and students from Leipzig). As well, there is very high consensus among the ratings of students from Bern and Leipzig (r = .992 [p < .001]).

[7] We also estimated a mixed model with random effects at the course and individual levels to take into account the fact that students are also nested in classes. Our findings remain the same if we control for

Munich SET questionnaire contains a self-generating panel identifier (e.g., Kearney et al. 1984; Yurek, Vasey, and Havens 2008). By asking students in the SET for a series of time constant, easily retrievable characteristics (sex, year of birth, degree of studies, the first two digits of their mother's birthday, the first two letters of their parents' first names, and the number of older brothers and sisters), we can link SET to individual students and are able to observe their rating and attendance behaviors for different courses over time, adding richness of information not available in other studies.

Furthermore, based on prior findings on determinants of SET ratings and attendance (e.g., Arulampalam, Naylor, and Smith 2012; Becker and Powers 2001; Bosshardt 2004; Devadoss and Foltz 1996; Wolbring 2012) we included control variables at the student, course, and department levels in our models (for details see table A1 in the appendix). Estimates for the effects of physical attractiveness on SET ratings and class attendance might be biased if courses of better-looking, younger instructors systematically differ in course characteristics (topic, difficulty, pace, time and day), and student composition (interest, motivation for course choice) from courses held by older, less attractive instructors.

Results

Teaching Quality. Physical attractiveness has a significant influence on SET in all models. The inclusion of control variables (model 2 in table 1) leads to the reduction of the effect that is now weaker than those reported in previous studies using SET data. For example, whereas Hamermesh and Parker (2005) find a maximum effect of roughly one out of five grade points, model 2 merely suggests a maximum effect of .037 times (10–1) equals .3 grade points. The

---

both forms of clustering, although standard errors are higher than in simple linear regression models. Standard errors are even more conservative in the fixed-effects models we report, making them our preferred results.

main reason for this difference is most probably the inclusion of student fixed effects and more extensive controlling for potential confounders.

[Table 1. Estimated Effects of Attractiveness and Intersexual Attraction]

Absenteeism from Class. The models for students' frequency of attendance show a similar picture. The physical attractiveness of the instructor is a weak but consistent predictor for the number of classes a student skipped. According to model 4, absenteeism is reduced by up to .7 classes per student and term if the instructor is physically very attractive instead of very unattractive, all else equal. Adjusting for the overall rating of teaching quality (model 5) and other controls (model 6) the effect decreases without disappearing. If students used physical attractiveness as an indicator for course quality then the relationship between the instructor's appearance and students' absenteeism should dissolve with the inclusion of these quality measures. This is clearly not the case. Since we control for students' perceptions of teaching quality in model 5 and 6, a preference for attractive individuals appears to be the most likely explanation for the significantly lower rates of absenteeism in classes held by good-looking instructors. Therefore, our results point to an attention boost for attractive instructors where students seem to derive utility from attending their classes and interacting with them.

Intersexual Attraction. Models 3 and 7 (table 1) test the proposition that attractiveness is especially influential if rater and ratee are of the opposite sex. As seen in figure 2 the results do not support this hypothesis. Physical attractiveness has a significant effect on students' ratings of teaching quality, which is comparable in strength for all three sex-constellations under consideration. The same holds true for the beauty effect on absenteeism, where the correlation is weakest if rater and ratee are of the same sex. Further analyses reveal that there is also no interaction between physical attractiveness and the sex of the instructor.

[Figure 2. Intersexual Attraction Effects on Course Ratings and Absenteeism]

Beauty Penalty. Next we ask whether the beauty premium turns into a penalty if attractive instructors do not meet students' expectations (i.e. blatant misconduct). In order to test this consideration we rely on three indicators: course was too difficult, pace of the course was too fast, and instructor did not adequately answer questions. Since theory only predicts an interaction for blatant misconduct, we generated a dummy variable which is 1 if misbehavior is likely strong (value of 5 on a rating scale from 1–5) and 0 otherwise. Only a very small fraction of all students indicate serious deficits regarding course difficulty (.9 percent), adequacy of instructor's answers (.9 percent), and course pace (2.3 percent). One consequence of this small amount of variation in the covariates is a low degree of statistical power in the following models.[8] Nonetheless, as seen in table 2 all three variables have strong and significant effects on SET. Ratings are .36–.94 grade points lower if the course was too difficult, the course pace was too fast, or answers were not satisfactory. This finding can be interpreted either as a cross-validation of SET as a measure of teaching quality or as the result of reciprocity. Moreover, for two of the three indicators the direction of the interaction with physical attractiveness is in line with our hypothesis of a beauty penalty. If students describe a course as too fast or the answers of the instructor as inadequate, physical attractiveness has a negative effect on SET which is comparable in size with the estimates for the beauty premium in the default case of no misconduct. Thus, although the interactions are not significant at conventional levels due to the low variance of the indicator variables, the effects appear to be substantial which partially supports the proposition of a beauty penalty.

[8] In addition, the wording of the items 'course difficulty' and 'course pace' was in the opposite direction of most items in the SET questionnaire. Since not all students might have recognized the reverse coding, measurement error could bias our estimates and reduce the power of our tests.

[Table 2. Estimated Effects of Attractiveness and Blatant Misconduct]

Summary and Discussion

The empirical findings support hypotheses 1 (attractiveness consensus), 2 (attractiveness attention boost), 3 (attractiveness treatment advantage), and partially hypothesis 5 (beauty penalty), but contradict hypothesis 4 (intersexual attraction).[9] Students agree in their attractiveness judgments, rate attractive instructors better in SET, punish them more severely for blatant misconduct, and attend their classes more frequently. Differences in productivity alone do not explain this last finding – students miss classes of attractive instructors less often even if we control for the teaching quality in students' perceptions. This is a central and new finding which adds to the literature on attractiveness and SET and is in line with the proposed mechanisms. However, in contrast to our theoretical considerations the effect of attractiveness does not vary with the sex of the instructor and the student. The association is not stronger if the student rater and the rated instructor are of the opposite sex raising questions about the theoretical mechanism at work.

[9] One could argue that intersexual attraction comes into effect only if the person rated is actually a potential mating partner. Especially, for female instructors above the fertile age this does not hold. Since no information on the age of the instructors is available we reran our analyses for midlevel faculty members only, but the results for this subsample still contradict the intersexual attraction hypothesis. Furthermore, one might object that at nearly all universities in the U.S., sexual contact between instructors and their current students is forbidden. This raises the question of whether any instructors are really viewed as "potential mating partners". Anecdotal evidence suggests that the campus actually is a market for potential mates. Moreover, although sexual contact between instructors and students is socially disapproved at the university under study it is not forbidden by the law or the university's rules of conduct.

Furthermore, although we provide a more demanding test for hypotheses on the effects of physical attractiveness than previous studies using SET data, some drawbacks remain. One major problem concerns the observational nature of the SET data making it impossible to rule out competing explanations for the effects of physical attractiveness. For example, excluding the possibility of confounders at the student level, the personality of attractive instructors could on average be more favorable and actually bring about better SET ratings and higher attendance rates. A second drawback is the inability to adequately test hypotheses on a beauty penalty. Since information on instructors' grading behavior was not available and since we could not systematically induce unfair treatment of students for ethical restrictions, we had to rely on murky indicators for blatant misconduct and utilize statistical tests with low power.

To tackle these potential shortcomings we conducted a series of laboratory experiments in which we control for instructor characteristics and teaching quality while manipulating the difficulty and timing of a test.

Study 2: Experimental approach

Study 2 uses data gathered in a series of laboratory experiments conducted at the University of Munich.[10] Participants were issued a CV (including a photograph) and listened to a short audio recording of a lecture which they were told was held by the person in the CV. They were also asked to complete a short test on this lecture and provide a rating for the lecture. The chronology of test and rating was systematically varied, providing a better test of the beauty penalty hypothesis than the non-experimental SET data.

[10] Part of those data were already used in Wolbring and Hellmann (2010). For this publication we extended the database and added the treatment 'easy test'.

Experimental Design and Research Strategy

The experiment included four dichotomous stimuli (physical attractiveness *high – low*, lecturer sex *male – female*, point in time of a test *prior to the evaluation – after the evaluation*, difficulty of test *hard – easy*). Participants were students of the University of Munich who were invited to participate in an experiment in the laboratory of the Institute of Sociology. In order to prevent bias, students from psychology and the social sciences were excluded from the sample, because they might have knowledge of research on attractiveness and SET. 323 participants showed up for the experiment (for the number of observations per treatment condition, see Table A2).

[Figure 3. Experimental Stimuli: Photos of the Lecturers]

The sequence of the different parts of the experiment in the laboratory was as follows: Participants were seated at work stations separated by blinds and received basic instructions. They were handed a CV and photograph of a fictitious male or female, attractive or unattractive "instructor" (see figure 3). The photographs were taken from the site http://www.beautycheck.de. This site offers digitally and systematically edited photographs spanning the spectrum from very unattractive to very attractive.

We chose our four pictures on the basis of the three criteria: First, the faces should clearly differ in beauty, but should only moderately deviate from average attractiveness.[11] In order to ensure an informed selection different photographs were evaluated by 11 male and 28 female students from the University of Munich on an 11-point attractiveness-scale (0 [very

[11] As shown in figure 3, faces differ on various dimensions of the construct "physical attractiveness", such as facial symmetry, weight, and skin texture. Since those different dimensions are varied simultaneously we can only identify the combined treatment effect of changes in physical attractiveness and not separate effects of each dimension in this experiment.

unattractive] – 10 [very attractive]). In line with the attractiveness consensus the students strongly agreed in their ratings (Cronbach's α = .97). As seen in Table A3 in the appendix photographs of unattractive males and females received attractiveness ratings of 3.6–3.8 from independent raters and similar ratings of 3.0–3.9 from the experimental subjects. As intended, attractive faces received considerably higher ratings by independent raters (male: 6.8; female: 7.4) but especially by the subjects (male: 7.4–8.0; female: 8.5). Second, we decided to show pictures of young instructors, since the mating argument from evolutionary theory only holds for fertile humans and since age difference is an important determinant of mate choice. Third, while varying physical attractiveness we tried to hold constant all other characteristics, such as age, clothes, hair length, and quality of the photograph in order to avoid effect heterogeneity and secure internal validity for the results.

After receiving the CV all participants listened to an eleven-minute audio recording of a lecture. The content of this lecture was a classical topic in the analysis of social structure: the discussion on classes, strata, and milieus, based on Hradil's (1987, pp. 7-11 and 72-74) textbook. Of course, the voice had to be altered alongside the sex on the fictitious CV. Instead of recording the lecture twice, with a female and male speaker, the voice was systematically adjusted for male or female voice characteristics using the software MorphVoxPro. This has the advantage of achieving maximum comparability between lectures by females and males concerning acoustic indicators of teaching quality such as voice melody and intonation.

After listening to this lecture, one half of the participants completed a slightly modified version of the standard SET questionnaire used at the Faculty of Social Sciences of the University Munich. Then either an easy or hard test on the content of the lecture followed.

The other half of the participants completed both documents in the opposite order.[12] In order to create a real incentive to take the test seriously and induce frustration in the hard test condition, the participants received a pay-off partly dependent on the number of correct answers. In this we assume that students interpret the difficulty of the test as intentional and consequential behavior on the part of the instructor. Such "spiteful acts" should induce reciprocal behavior on the part of the students leading to an especially strong penalty for attractive instructors in the case of the overly difficult test. Every participant received a fixed show-up fee of EUR 4.00 and, in addition, EUR 0.50 for every correct answer in the test. On average, participants assigned to the hard (easy) test answered 1.90 (6.90) questions correctly and earned EUR 4.95 (EUR 7.45). Participants were paid at the end of the 30-minute experiment, but received feedback on their performance (grade, number of correct answers) immediately after the completion of the test.

The physical attractiveness of the instructors (as depicted on the CV) is the central explanatory variable in our analyses. We include three dummy variables, attractive male, attractive female, and unattractive male, in the following models which leaves unattractive female as the reference category.

Similar to study 1 the outcome of interest is students' ratings of the overall quality of the lecture ("All in all, how do you rate the overall quality of this lecture (content, style of presentation)?" very good [5.0] – insufficient [1.0]), which we analyze using simple linear regressions. In a first step, we examine only the ratings of those students who filled out the questionnaire before completing the test to show effects of physical attractiveness and intersexual attraction. Then, we additionally focus on those ratings completed after the test

[12] English versions of these documents, i.e. the SET questionnaire as well as the test are available on http://www.[redacted].

exploring whether the strength and direction of the beauty bias depends on the difficulty and timing of the test.

Results

Physical Attractiveness. The effect of instructors' looks on SET is similar in size to that reported in study 1. On average, attractive instructors received ratings which are .24 grade points higher (not tabulated). The beauty premium is markedly stronger for female instructors (.40) than for male instructors (.10, difference between the effects "attractive male" and "unattractive male") and significant at the 10 percent level for the former (model 1 in table 3). Furthermore, physical attractiveness is significantly related to performance in the hard test with students listening to an attractive instructor answering .44 more questions correctly (two-tailed t-test: $p = .025$; $N = 149$). However, there are no differences between the two treatment groups in the performance in the easy test (two-tailed t-test: $p = .800$; $N = 176$). Controlling for the number of correct test answers in model 2 the effect of physical attractiveness on SET remains remarkably stable. Because of this finding and since all students listened to exactly the same lecture we can rule out the alternative explanation that the quality of the talk and learning success cause the beauty premium.

Beauty Penalty. Surprisingly, the positive effect of physical attractiveness disappears for female instructors and even becomes negative for male instructors when we look at ratings given after the test (model 3). This finding could be related to the fact that students react to the test in a reciprocal way and instructors are punished for demanding too much. Generally, individual performance in the test is significantly related to students' ratings. In addition, students who completed a hard instead of an easy test rate the teaching significantly worse

(analysis not reported).[13] Thus, if students have higher expectations about the behavior of physically attractive instructors and if they sanction them more severely for failures the attractiveness effects could be masked in model 3. This consideration is supported by the results of model 4 where we added interaction terms between attractiveness and the timing and difficulty of the test. Physical attractiveness has a weak positive effect on students' ratings as long as the test takes place after the SET.[14] This is also demonstrated in figure 4, where we depict the change in ratings compared to the default case that SET are conducted before the tests, the same holds if the test is easy. However, if attractive instructors demand too much from their students and the test takes place before the SET the premium turns into a penalty. This means that attractive instructors who gave students a hard test were punished more severely. Although the interaction effects "attractive male instructor hard test prior SET" and "attractive female instructor hard test prior SET" are similar in strength (both -.51) the beauty penalty is especially strong for male instructors. This is because female instructors receive a beauty premium of .23 that partly offsets their penalty (see model 4).

[Figure 4. Beauty Effects and Test Difficulty]

[Table 3. Effects of Attractiveness and Interaction with the Timing and Difficulty of a Test]

Intersexual Attraction. Finally, we estimated models 2 and 4 separately for the cases where students and instructors are of the opposite sex. As in study 1 the effects of physical attractiveness for females rating males and males rating females do not significantly differ

[13] Due to multicollinearity we were not able to simultaneously include the number of correct test answers and the treatment dummy for test difficulty in the models (r = .76 for number of correct answers in the hard test; r =.96 for number of correct answers in the easy tests).

[14] It is worth noting that before listening to the lecture all participants knew a test was to be completed during the experiment.

from the average effect if the SET took place before the test (see models 2a and 2b). Looking at the analyses for the beauty penalty we find pretty similar results (see models 4a and 4b): The main effects of beauty do not substantially change if rater and ratee are of the opposite sex. The beauty penalty is also quite similar in size if females rate males and becomes slightly stronger if males rate females. However, there are no statistically significant differences between the penalty effects for different sex-combinations, although missing significance should not be overstated in this case, since the groups become rather small in case numbers.

Summary and Discussion

The empirical findings support hypotheses 1 (attractiveness consensus), 3 (attractiveness treatment advantage), and 5 (beauty penalty), but contradict 4 (intersexual attraction). Although controlling for teaching quality and students' test performance we still find effects of attractiveness. It is remarkable that the beauty effects found in this study are very similar in size to those in study 1 given that we tackled the questions with two very different research strategies. While the advantage of study 1 lies in the external validity of the finding, we were able to manipulate and thus isolate the effect of attractiveness in study 2. According to our findings the beauty premium is markedly higher for female instructors. Another central finding of this study is that especially male, but also female instructors are punished more severely for failures if they are good-looking. The instructors' appearance obviously influences students' aspiration levels and results in differential treatments. Whether attractive instructors are treated better or worse depends on their behavior. The more they deviate from students' expectations, the more likely it is that the attractiveness effect becomes negative.

Thus, although the data base is not very large, these experiments lead to promising results on the effects of attractiveness and the processes at hand. Our approach appears to be a fruitful avenue for future research, especially since certain questions remain unanswered. First, it is unclear at what point the beauty premium turns into a penalty. In this experiment we induced

the beauty penalty using a difficult test administered to the student subjects. In contrast, students given an easy test did not provoke similar sanctions. However, we cannot say how difficult the test has to be to reverse the positive beauty effect.

Second, we also do not know whether students would react similarly to modified versions of the experimental stimuli. On the one hand this concerns additional failure treatments such as bad presentation skills, lack of competence (e.g., teaching obviously incorrect information), or impolite behavior towards the students. On the other hand one might ask whether the findings are robust given variations in characteristics of the instructors shown on the photographs, such as age, clothing, and hair style, which might moderate the beauty premium and penalty. For example, one could hypothesize that gender-specific attractiveness stereotypes are different for young and old people, or for males and females with long and short hair, or for those who wear more professional, fashionable or sexually revealing clothing. However, the required number of cases needed to identify the effects of several of these potential moderator variables grows exponentially, making this goal practically impossible in a single experiment leaving alternative potential moderators open for future research.

Third, since human perception is relative, effects of physical attractiveness might not only depend on an individual's appearance, but also on the looks of people surrounding her. Such a frog pond hypothesis as originally proposed by Davis (1966; see also Frank 1985) has seldom been tested empirically, but could be explored with the proposed experimental approach (for a test see Wolbring and Hellmann 2010).

Conclusions

In this paper we examined different effects of attractiveness in everyday life drawing on the empirical example of SET. To clarify how beauty works we linked and contrasted two theoretical perspectives, evolutionary theory and expectation states theory, and extended the

argument to the case of non-negligible expectation-disconfirming behavior. Although the analytical foci and "stopping rules" are very different both perspectives agree in their general implications for the problem at hand. After illustrating the precise structure of the proposed mechanisms in DAGs, we tackled common problems of causality in studies on the effects of physical attractiveness by using two distinct empirical strategies: In study 1 we drew on SET collected at the University of Munich and were able to link publically available photographs of teachers and repeated observations of students over time to the data, giving us the possibility to control for many confounders. Besides results on attractiveness effects on SET ratings we provided evidence for an attention boost while controlling teaching quality in students' perceptions. In study 2, we utilized an experimental setup to examine the effects of beauty and fair or unfair treatment of students by their instructors. The experimental approach allowed us to hold all other aspects of teaching quality constant, enabling us to identify interactions between the difficulty of a test and instructors' physical attractiveness.

Combining the strengths of both methodologies has proven fruitful. It is remarkable that these different empirical strategies find consistent effects of attractiveness. Raters revealed a high degree of consensus when evaluating instructors' appearances and these beauty evaluations in turn resulted in differential treatments of instructors. Thereby, differences in instructors' productivity could not solely explain the latter effect. The results from study 1 showed that courses of attractive instructors were attended more regularly even if we controlled for students' perceptions of teaching quality. Analyses of the experimental data in study 2 revealed differences in the SET ratings of attractive and unattractive lecturers although students listened to exactly the *same* audio recording and performed similarly in a test under different treatment conditions. We conclude from these findings that physical attractiveness is an important, but widely neglected facet of social inequality which deserves more attention in sociological research.

Moreover, our findings draw attention to another blind spot of sociological theory. Expectation states theory proposes that individuals rely on diffuse status characteristics to form expectations about traits and behavior of others in social interactions. Thereby, the burden of proof lies on demonstrating that a given status characteristic is *not* a valid signal for traits of interest: if not clearly proved otherwise, status is always relevant in the situation at hand and can even blind people into attributing failures of attractive individuals to external sources. Thus, due to positive reinforcement and glamour effects expectations usually are self-confirmatory and stabilize over time. However, the framework leaves out the important question of what happens in a situation in which there is overwhelming non-negligible proof that subjects do not meet the expected qualities. Our findings suggest that such an obvious lack of fit between expectations and observed behavior turns the status premium into a penalty: instructors were sanctioned for administering a hard test more severely if they were good-looking. Thereby, we expect this effect to hold only if the observed behavior is interpreted as an act of intentional friendliness or hostility and directly affects the onlookers. Future research should integrate this lack-of-fit-argument into expectation states theory and to empirically determine the scope conditions of such inverse status effects.

Furthermore, the hypothesis of intersexual attraction was not corroborated by our data. Effects of physical attractiveness were not significantly stronger if rater and ratee were of different sex in either study. We therefore conclude that evolutionary considerations on mating cannot be easily transmitted to topics which have little relation to survival and reproduction (Mulford et al. 1998) – as is the case with SET. Students might regard physical attractiveness as a valid signal for teaching ability irrespective of the combination of their own and the instructor's sex. This is in line with a general human preference for other good-looking objects and hence does not render the evolutionary framework worthless in beauty research.

An alternative argument could be that the effects of beauty are the result of cultural stereotypes and prejudices toward physically attractive and unattractive people, and are totally unrelated to evolutionary processes. Expectation states theory allows an analysis of beauty effects independent from any considerations of the reasons for the variability in physical attractiveness. As well, the results which indicate the absence of any cross-gender-effects are in line with those arguments from expectation states theory which state that diffuse status signals combine additively. Furthermore, our experimental results corroborate considerations that imply an interaction of attractiveness and gender, and more specifically stronger effects of beauty for female instructors.

Expectation states theory offers proximate causes for the formation and stabilization of expectations, such as path dependencies, self-fulfilling prophecies, opinion leadership, and media influences. However, this leaves the question unanswered why specific stereotypes and prejudices are the way they are. The evolutionary framework adds to this explanatory problem hinting at ultimate causes for the emergence and general form of expectations. While expectation states theory does not specify the origins of differences in ability between status groups evolutionary theory offers a deeper underpinning for the signaling value of beauty. Hence, as Bearman (2008: SV/SVI) has previously emphasized: “In the most positive vein, the current interest in genetics may provide a new lever for sociologists to escape what appears at times to be a hegemonic focus on just "our" society”. We extend this to the value of being able to answer the question how partially similar beauty standards could have emerged in different cultures and why it might be insufficient to assert they may have come about by pure chance. However, evolutionary theory does not elaborate on self-enforcing dynamics of social interactions. Expectation states theory can open this black box by offering a better understanding for the social processes at hand, since evolutionary influences are mediated by

social conditions. Thus, although their foci are very different, expectation states and evolutionary theory can profit from each other through further interlinkage and exchange.

Last but not least, this research does not only help us explore theory, but it also helps us understand a major part of our academic lives: teaching. SET research is so popular because it allows us to understand the unique world we inhabit just a little bit better. One central take away from this study is that social mechanisms which are well-documented for social interactions in everyday life, such as stereotyping, discrimination, and reciprocity, also operate in the classroom. SET are not context-free and, thus, should not be automatically linked to incentive schemes in higher education. Although numerous studies suggest that SET measures are valid measures of teaching quality *on average* (Marsh 2007; Spooren, Brockx, and Mortelmans 2013), extraneous influences which are unrelated to teaching performance, such as the beauty premium, can bias ratings of *individual instructors*. Hence, increasing the stakes of SET might not only foster manipulations of the ratings by means of easy tests, lenient grading, and other "presents" to the students (e.g., finishing class early, reducing the amount of assignments, distributing candy for in-class tasks), but might even have the unintended consequence of discouraging instructors who feel treated unfairly.

Appendix

Table A1. Descriptive Statistics for Variables Used in Study 1

Variable	Definition	Mean	SD
Overall rating: teaching quality	"On a continuous grading scale from excellent [5.0] to insufficient [1.0]: How do you rate this course?	1.993	.679
Number of classes skipped	"Up to now, how many classes of this course did you miss?"	1.224	1.265
Physical attractiveness	Average attractiveness of the face evaluated by independent students raters: very unattractive [0] – very attractive [10]	4.516	1.031
Sex of the instructor	Female [1]; male [0]	.241	.428
Sex of the student rater	Female [1]; male [0]	.640	.480
Performance record	"Do you need a performance record?" yes [1]; no [0]	.909	.288
Course size	Number of students who completed an SET form in the course	88.594	104.337
Department	Department which offered the course (reference category: communication science)		
Political Science		.329	.470
Sociology		.360	.480
Course day	Weekday of the course (reference category: Monday)		
Tuesday		.227	.419
Wednesday		.270	.444
Thursday		.226	.418
Friday		.025	.156
Course time	Starting time of the course (reference category: 8)		
10 a.m.		.359	.480
12 p.m.		.171	.376
2 p.m.		.158	.365
4 p.m.		.126	.331
6 p.m. and later		.063	.244
Summer term	Equals 1 if the course is held during a summer term, otherwise winter term	.500	.500
Course pace	"I could not follow the pace of the course." totally agree [1] – totally disagree [5]	3.906	1.031
Course difficulty	"The course was too difficult" totally agree [1] – totally disagree [5]	4.124	.919
Prior interest	"I chose the course, because I was interested in its content" totally agree [1] – totally disagree [5]	2.427	1.256
Preparation for the course	"On average, how many minutes per week did you prepare for the course?"	56.624	58.293
Courseload	"How many hours per week do you take for credit in this semester?"	17.058	5.651
Workload	"How many hours per week do you work for payment in this semester?"	8.218	8.155
Semester of study	Student's subject-related semester of study	3.669	2.480

Note: Means and standard deviations observations used in model 1 and 4 in table 1 (N = 10159).

Table A2. Number of Observations by Treatment Condition

		SET before test		SET after test		
		easy test	**hard test**	**easy test**	**hard test**	
attractive	**male instructor**	26	19	21	20	**86**
	female instructor	20	16	22	20	**78**
unattractive	**male instructor**	24	19	21	19	**83**
	female instructor	20	20	22	16	**78**
		90	**74**	**86**	**75**	

Table A3. Ratings of Photographs by Independent Raters and Experimental Subjects

Picture	Male; unattractive	Male; attractive	Female; unattractive	Female; attractive
Independent Rater	3.58	6.82	3.77	7.44
Subjects with Easy Test	3.91	7.96	3.93	8.50
Subjects with Hard Test	3.92	7.41	2.97	8.53

Note: Rating scale for attractiveness ranges from 0[–] to 10[+].

References

Adloff, Frank and Steffen Mau. 2006. "Giving Social Ties. Reciprocity in Modern Society." *European Journal of Sociology* 47(1):93–123.

Allison, Paul. D. 2009. *Fixed Effects Regression Models.* Thousand Oaks, CA: Sage.

Ambady, Nalini and Robert Rosenthal. 1993. "Half a Minute: Predicting Teacher Evaluations From Thin Slices of Nonverbal Behavior and Physical Attractiveness." *Journal of Personality and Social Psychology* 64(3):431–41.

Andreoni, James and Ragan Petrie. 2008. "Beauty, Gender and Stereotypes: Evidence from Laboratory Experiments." *Journal of Economic Psychology* 29(1):73–93.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton University Press.

Arulampalam, Wiji, Robin A. Naylor and Jeremy Smith. 2012. "Am I Missing Something? The Effects of Absence from Class on Student Performance." *Economics of Education Review* 31(4):363–75.

Axelrod, Robert. 1984. *The Evolution of Cooperation.* New York: Basic Books.

Axelrod, Robert. 1997. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration.* Princeton, NJ: Princeton University Press.

Barkow, Jerome H., Leda Cosmides and John Tooby. 1992. *The Adapted Mind. Evolutionary Psychology and the Generation of Culture.* New York: Oxford University Press.

Bassili, John N. 1981. "The Attractiveness Stereotype: Goodness or Glamour?" *Basic and Applied Social Psychology* 2(4):235–52.

Bearman, Peter. 2008. "Introduction. Exploring Genetics and Social Structure." *American Journal of Sociology* 114:SV–SX.

Becker, William E. and John R. Powers. 2001. ''Student Performance, Attrition, and Class Size Given Missing Student Data.'' *Economics of Education Review* 20(4):377–88.

Belot, Michèle, V. Bhaskar, and Jeroen de Ven. 2012. "Beauty and the Sources of Discrimination." *Journal of Human Resources* 47(3):851–72.

Berger, Joseph, Bernard P. Cohen, and Morris Zelditch, Jr. 1972. "Status Characteristics and Social Interaction." American Sociological Review 37(3):241–55.

Berger, Joseph and M. Hamit Fiske. 1972. "Diffuse Status Characteristics and the Spread of Status Value: A Formal Theory." American Journal of Sociology 111(4):1038–79.

Berger, Joseph, M. Hamit Fiske, Robert Z. Norman, and Morris Zelditch, Jr. 1977. *Status Characteristics and Social Interaction.* New York: Elsevier.

Biddle, Jeff E. and Daniel S. Hamermesh. 1998. "Beauty, Productivity and Discrimination; Lawyers' Looks and Lucre." *Journal of Labor Economics* 16(1):172–201.

Blau, Peter M. 1964. *Exchange and Power in Social Life.* New York: Wiley.

Bosshardt, William. 2004. "Student Drops and Failure in Principles Courses." *Journal of Economic Education* 35(2):111–28.

Bowles, Samuel and Herbert Gintis. 2011. *A Cooperative Species. Human Reciprocity and its Evolution.* Princeton, NJ: Princeton University Press.

Buck, Stephen and Drew Tiene. 1989. "The Impact of Physical Attractiveness, Gender, and Teaching Philosophy on Teacher Evaluations." *Journal of Educational Research* 82(3):172–7.

Buss, David M. 1989. "Sex Differences in Human Mate Preferences: Evolutionary Hypotheses Tested in 37 Cultures." *Behavioral and Brain Sciences* 12(1):1–49.

Buss, David M. 1999. *Evolutionary Psychology. The New Science of the Mind.* Boston, MA: Allyn & Bacon.

Buss, David M., ed. 2005. *The Handbook of Evolutionary Psychology.* Hoboken, NJ: Wiley.

Chase, Ivan D. and W. Brent Lindquist. 2009. "Dominance Hierarchies" Pp. 566–93 in *The Oxford Handbook of Analytical Sociology*, edited by P. Hedström and P. Bearman. Oxford, UK: Oxford University Press.

Cosmides, Leda and John Tooby. 1989. "Evolutionary Psychology and the Generation of Culture, Part II: Case Study: A Computational Theory of Social Exchange." *Ethology and Sociobiology* 10(1):51–97.

Currie, Thomas E. and Anthony C. Little. 2009. "The Relative Importance of the Face and Body in Judgments of Human Physical Attractiveness." *Evolution & Human Behavior* 30(6):409–16.

Dabbs, James M., Jr. and Neil A. Stokes III. 1975. "Beauty Is Power: The Use of Space on the Sidewalk." *Sociometry* 38(4):551–7.

Darley, John M. and Russel H. Fazio. 1980. "Expectancy Confirmation Processes Arising in the Social Interaction Sequence." *American Psychologist* 35(10):867–81.

Davis, James A. 1966. "The Campus as a Frog Pond: An Application of the Theory of Relative Deprivation to Career Decisions of College Men." *American Journal of Sociology* 72(1):17–31.

Devadoss, Stephen and John Foltz. 1996. "Evaluation of Factors Influencing Student Class Attendance and Performance." *American Journal of Agricultural Economics* 78:499–507.

Dion, Karen, Ellen Berscheid, and Elaine Walster. 1972. "What Is Beautiful Is Good." *Journal of Personality and Social Psychology* 24(3):285–90.

DiPrete, Thomas and Gregory M. Eirich. 2006. "Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments." *Annual Review of Sociology* 32:271–97.

Dutton, Denis. 2009. *The Art Instinct. Beauty, Pleasure, and Human Evolution.* New York: Bloomsberry Press.

Eagly, Alice H., Richard D. Ashmore, Mona G. Makhijani, and Laura C. Longo. 1991. "What is Beautiful Is Good, But. . . : A Meta-Analytic Review of Research on the Physical Attractiveness Stereotype." *Psychological Bulletin* 110(1):109–28.

Elder, Glen H., Jr. 2008. "Appearance and Education in Marriage Mobility." *American Sociological Review* 34(4):519–33.

Elwert, Felix. 2013. "Graphical Causal Models." Pp. 245-274 in *Handbook of Causal Analysis for Social Research* edited by Stephen L. Morgan. Dorddrecht, NL: Springer.

Felton, James, John Mitchell, and Michael Stinson. 2004. "Web-based Student Evaluations of Professors: The Relation between Perceived Quality, Easiness and Sexiness." *Assessment & Evaluation in Higher Education* 29(1):91–108.

Felton, James, Peter T. Koper, John Mitchell, and Michael Stinson. 2008. "Attractiveness, Easiness, and Other Issues: Student Evaluations of Professors on RateMyProfessors.com." *Assessment & Evaluation in Higher Education* 33(1):45–61.

Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.

Frank, Robert H. 1985. *Choosing the Right Pond. Human Behavior and the Quest for Status*. New York: Oxford University Press.

Freese, Jeremy. 2000. *What Should Sociology Do About Darwin? Evaluating Some Potential Contributions of Sociobiology and Evolutionary Psychology to Sociology.* PhD Dissertation, Department of Sociology, Indiana University. Retrieved December 4, 2012 (http://jeremyfreese.com/docs/freese_dissertation.pdf).

Freese, Jeremy. 2008. "Genetics and the Social Science Explanation of Individual Outcomes." *American Journal of Sociology* 114:S1–S35.

Freese, Jeremy and Sara Shostak. 2009. "Genetics and Social Inquiry." *Annual Reviews of Sociology* 35:107–28.

Gangestad, Steven W. and Glenn J. Scheyd. 2005. "The Evolution of Physical Attractiveness." *Annual Review of Anthropology* 34:523–48.

Gigerenzer, Gerd. 2008. *Rationality for Mortals. How People Cope with Uncertainty.* Oxford, UK: Oxford University Press.

Gillen, Barry. 1981. "Physical Attractiveness: A Determinant of Two Types of Goodness." *Personality and Social Psychology Bulletin* 7(2):277–281.

Goebel, Barbara L. and Valjean M. Cashen. 1979. “Age, Sex, and Attractiveness as Factors in Student Ratings of Teachers: A Developemental Study.” *Journal of Educational Psychology* 71(5):646–53.

Gouldner, Alvin W. 1960. “The Norm of Reciprocity. A Preliminary Statement.” *American Sociological Review* 25(2):161–78.

Grammer, Karl, Bernhard Fink, Anders P. Møller and Randy Thornhill. 2003. “Darwinian Aesthetics: Sexual Selection and the Biology of Beauty.” *Biological Reviews of the Cambridge Philosophical Society* 78(3):385–407.

Guo, Gang. 2006. “The Linking of Sociology and Biology.” *Social Forces* 85(1):145–49.

Hamermesh, Daniel S. 2006. “Changing Looks and Changing ‘Discrimination’: The Beauty of Economists.” *Economics Letters* 93(3):405–12.

Hamermesh, Daniel S. 2011. *Beauty Pays.* Princeton, NJ: Princeton University Press.

Hamermesh, Daniel S. and Jeff E. Biddle. 1994. “Beauty and the Labor Market.” *American Economic Review* 84(5):1174–94.

Hamermesh, Daniel S. and Amy M. Parker. 2005. “Beauty in the Classroom. Instructors’ Pulchritude and Putative Pedagogical Productivity.” *Economics of Education Review* 24(4):369–76.

Hamermesh, Daniel S., Xing Meng, and Junsen Zhang. 2002. “Dress for Success – Does Primping Pay?” *Labour Economics* 9(3): 361–73.

Heilman, Madeline and Tyler G. Okimoto. 2007. “Why Are Women Penalized for Success at Male Tasks?: The Implied Communality Deficit.” *Journal of Applied Psycholoy* 92(1):81–92.

Heilman, Madeline and Lois R. Saruwatari. 1979. “When Beauty is Beastly: The Effects of Appearance and Sex on Evaluations of Job Applicants for Managerial and Nonmanagerial Jobs.” *Organizational Behavior and Human Performance* 23(4):360–72.

Henss, R. 1992. *Spieglein, Spieglein an der Wand. . . . Geschlecht, Alter und physische Attraktivität.* Weinheim, Germany: Psychologie Verlags Union.

Hopcroft, Rosemary L. 2009. “The Evolved Actor in Sociology.” *Sociological Theory* 27(4):390–406.

Hysom, Stuart J. 2009. “Status Valued Goal Objects and Performance Expectations.” *Social Forces* 87(3): 1623–48.

Jackson, Linda A., John E. Hunter, and Carole N. Hodge. 1995. “Physical Attractiveness and Intellectual Competence: A Meta-Analytic Review.” *Social Psychology Quarterly* 58(2):108–22.

Jæger, Mads M. 2011. “‘‘A Thing of Beauty is a Joy Forever?’’ Returns to Physical Attractiveness over the Life Course.” Social Forces 89(3):983–1004.

Kanazawa, Satoshi. 2001. “De Gustibus Est Disputandum.” *Social Forces* 79(3):1131–63.

Kanazawa, Satoshi. 2011. “Intelligence and Physical Attractiveness.” *Intelligence* 39(1):7–14.

Kearney, Kathleen A., Ronald H. Hopkins, Armand L. Mauss, and Ralph A. Weisheit. 1984. “Self-Generated Identification Codes for Anonymous Collection of Longitudinal Questionnaire Data.” *Public Opinion Quarterly* 48(1B):370–78.

Klein, Markus and Ulrich Rosar. 2006. “Das Auge hört mit! Der Einfluss der physischen Attraktivität des Lehrpersonals auf die studentische Evaluation von Lehrveranstaltungen – eine empirische Analyse am Beispiel der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln.” *Zeitschrift für Soziologie* 35(4):305–16.

Kolm, Serge-Christophe and Jean M. Ythier. 2006. *Handbook of the Economics of Giving, Altruism,and Reciprocity. Vol. 1.* Oxford, UK: Elsevier.

Langlois, Judith H., Lisa Kalakanis, Adam J. Rubenstein, Andrea Larson, Monica Hallam, and Monica Smoot. 2000. “Maxims or Myths of Beauty? A Meta-Analytical and Theoretical Review.” *Psycholgical Bulletin* 126(3):390–423.

Langlois, Judith H. and Lori A. Roggman. 1990. “Attractive Faces Are Only Average.” *Psychological Science* 1(2):115–21.

Langlois, Judith H., Lori A. Roggman and Loretta A. Rieser-Danner. 1990. "Infants' Differential Social Responses to Attractive and Unattractive faces." *Developmental Psychology* 26(1):153–59.

Lenski, Gerhard. 2005. *Ecological Evolutionary Theory. Principles and Applications.* London, UK: Paradigm Publishers.

Lopreato, Joseph and Timothy Crippen. 2002. *Crisis in Sociology: The Need or Darwin.* London, UK: Transaction Publishers.

Malinowski, Bronislaw. 1922. *Argonauts of the Western Pacific.* London, UK: Routledge & Kegan Paul.

Maner, Jon K., Matthew T. Gailliot, D. Aaron Rouby, and Saul L. Miller. 2007. "Can't Take My Eyes off You: Attentional Adhesion to Mates and Rivals." *Journal of Personality and Social Psychology* 93(3):389–401.

Margolin, Leslie and Lynn White. 1987. "The Continuing Role of Physical Attractiveness in Marriage." *Journal of Marriage and the Family* 49(1):21–7.

Marsh, Herbert W. 2007." Students' Evaluations of University Teaching: A Multidimensional Perspective." Pp. 319-84 in *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective,* edited by R. P. Perry and J C. Smart. New York: Springer.

Mauss, Marcel. [1923] 1990. *Die Gabe. Form und Funktion des Austauschs in archaischen Gesellschaften.* Frankfurt am Main, Germany: Suhrkamp.

Mayr, Ernst. 2001. *What Evolution Is.* New York: Basic Books.

McClintock, Elizabeth A. 2014. "Beauty and Status: The Illusion of Exchange in Partner Selection? *American Sociological Review,* doi: 10.1177/0003122414536391.

Miller, Geoffrey F. 2001. *The Mating Mind. How Sexual Choice Shaped the Evolution of Human Nature*. New York: Anchor Books.

Miller, Geoffrey F. and Peter M. Todd. 1998. "Mate Choice Turns Cognitive." *Trends in Cognitive Sciences* 2(5):190–8.

Miller, Richard W. 1987. *Fact and Method: Explanation, Confirmation and Reality in the Natural and Social Sciences*. Oxford: Princeton University Press.

Mobius, Markus M. and Tanya S. Rosenblat. 2006. "Why Beauty Matters." *American Economic Review* 96(1):222–35.

Mocan, Naci and Erdal Tekin. 2010. "Ugly Criminals." *The Review of Economics and Statistics* 92(1):15–30.

Møller, Anders Pape and John P. Swaddle. 1997. *Asymmetry, Developmental Stability, and Evolution.* Oxford, UK: Oxford University Press.

Morgan, Stephen L. and Christopher Winship. 2007. *Counterfactual and Causal Inference. Methods and Principles for Social Research.* Cambridge: Cambridge University Press.

Morrow, Paula C. 1990. "Physical Attractiveness and Selection Decision Making." *Journal of Management* 16(1):45–60.

Mueller, Ulrich and Allan Mazur. 1996. „Facial Dominance of West Point Cadets as a Predictor of Later Military Rank." *Social Forces* 74:823–50.

Mulford, Matthew, John Orbell, Catherine Shatto, and Jean Stockard. 1998. "Physical Attractiveness, Opportunity, and Success in Everyday Exchange." *American Journal of Sociology* 103(6):1565–92.

Nielsen, François. 1994. "Sociobiology and Sociology." *Annual Review of Sociology* 20:267–303.

Patzer, Gordon L. 1985. *The Physical Attractiveness Phenomena.* New York: PlenumPress.

Patzer, Gordon L. 2007. *Why Physically Attractive People are More Successful. The Scientific Explanation, Social Consequences, and Ethical Problems.* New York: Edwin Mellen Press.

Pearl, Judea. [2000] 2009. *Causality: Models, Reasoning, and Inference* (2nd edition). Cambridge, UK: Cambridge University Press.

Peters, Marianne, Gillian Rhodes, and Leigh W. Simmons. 2007. "Contributions of the Face and Body to Overall Attractiveness." *Animal Behavior* 73:937-942.

Pinker, Steven. 2003. *The Modern Denial of Human Nature.* Reprint Edition. London, UK: Penguin Books.

Rhodes, Gillian. 2006. “The Evolution of Facial Attractiveness.” *Annual Review of Psychology* 57:199–226.

Rhodes, Gillian, Fiona Proffitt, Jonathon M. Grady, and Alex Sumich. 1998. “Facial Symmetry and the Perception of Beauty.” *Psychonomic Bulletin & Review* 5(4):659–669.

Rhodes, Gillian, Leslie A. Zebrowitz, Alison Clark, S. Michael Kalick, Amy Hightower, and Ryan McKay. 2001. “Do Facial Averageness and Symmetry Signal Health?” *Evolution and Human Behavior* 22(1):31–46.

Rhodes, Gillian, and Leslie A. Zebrowitz (Ed.). 2002. *Facial Attractiveness Evolutionary, Cognitive, and Social Perspectives.* Westpoint: Ablex.

Richerson, Peter J. and Robert Boyd. 2004. *Not by Genes Alone. How Culture Transformed Human Evolution*. Chicago/London, Uk: University of Chicago Press.

Ridgeway, Cecilia L. 1991. “The Social Construction of Status Value: Gender and Other Nominal Characteristics.” *Social Forces* 70(2):367–380.

Ridgeway, Cecilia L. 2011. *Framed by Gender. How Gender Inequality Persists in the Modern World.* Oxford, UK: Oxford University Press.

Ridgeway, Cecilia L. Kirsten Backor, Yan E. Li, Justine E. Tinkler, and Kristan G. Erickson. 2009. “How Easily Does a Social Difference Become a Status Distinction: Gender Matters.” *American Sociological Review* 74(1):44–62.

Ridgeway, Cecilia L. and Lynn Smith-Lovin. 1999. “The Gender System and Interaction.” *Annual Review of Sociology* 25:191–216.

Riniolo, Todd C., Katherine C. Johnson, Tracy R. Sherman, and Julie Misso. 2006. “Hot or Not: Do Professors Perceived as Physically Attractive Receive Higher Student Evaluations?” *Journal of General Psychology* 133(1):19–35.

Ritts, Vicki, Miles L. Patterson and Mark E. Tubbs. 1992. “Expectations, Impressions, and Judgments of Physically Attractive Students – A Review.” *Review of Educational Research* 62(4):413–26.

Rosar, Ulrich 2009. *Physische Attraktivität und soziale Ungleichheit. Ein Forschungsprogramm.* Habilitation thesis: University of Cologne.

Rosar, Ulrich, Tilo Beckers and Markus Klein. 2008. “The Frog Pond Beauty Contest. Physical Attractiveness and Electoral Successof the Constituency Candidates at the North-Rhine-Westphalia State Election of 2005.” *European Journal of Political Research* 47(1):621–45.

Rosar, Ulrich, Tilo Beckers, and Markus Klein. 2012. “Magic Mayors. Predicting Electoral Success from Candidates‘ Physical Attractiveness under the Conditions of a Presidential Electoral System.” *German Politics* 21(4):372–91.

Rosar, Ulrich and Markus Klein. 2009. „Mein(schöner)Prof.de. Die physische Attraktivität des akademischen Lehrpersonals und ihr Einfluss auf die Ergebnisse studentischer Lehrevaluationen.“ *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 61(4):621–45.

Samuels, Curtis A. and Richard Ewy. 1985. “Aesthetic Perception of Faces During Infancy.” *British Journal of Developmental Psychology* 3(3):221–228.

Schulmann, Steven R. 1978. „Kin Selection, Reciprocal Altruism, and the Principle of Maximization: A Reply to Sahlins.“ *Quarterly Review of Biology* 53(3):283–6.

Sigall, Harold and Nancy Ostrove. 1975. “Beautiful but Dangerous: Effects of Offender Attractiveness and Nature of the Crime on Juridic Judgment.” *Journal of Personality and Social Psychology* 31(3):410-414.

Simmel, Georg. 1900. *Philosophie des Geldes.* Leipzig, Germany: Duncker und Humblot.

Simmel, Georg. 1908. *Soziologie. Untersuchungen über die Formen der Vergesellschaftung.* Leipzig, Germany: Duncker und Humblot.

Simpson, Jeffry A., Steven W. Gangestad and Margaret Lermer. 1990. “Perception of Physical Attractiveness: Mechanisms Involved in the Maintenance of Romantic Relationships.” *Journal of Personality and Social Psychology* 59(6):1192–201.

Simon, Herbert A. 1957. *Models of Man.* New York: Wiley.

Snyder, Mark, Ellen Berscheid, and Alana Matwychuk. 1988. "Orientations Towards Personnel Selection: Differential Reliance on Appearance and Personality." *Journal of Personality and Social Psychology* 54(6): 972–79.

Solnick, Sara J. and Maurice E. Schweitzer. 1999. "The Influence of Physical Attractiveness and Gender on Ultimatum Game Decisions." *Organizational Behavior and Human Decision Processes* 79(3):199–215.

Spooren Pieter , Bert Brockx, and Dimitri Mortelmans. 2013. "On the Validity of Student Evaluation of Reaching: the State of the Art." *Review of Educational Research* 83(4):598–642.

Süssmuth, Bernd. 2006. "Beauty in the Classroom: Are German Students Less Blinded? Putative Pedagogical Productivity Due to Professors' Pulchritude: Peculiar or Pervasive?" *Applied Economics* 38(2):231–8.

Symons, Donald (1979): *The Evolution of Human Sexuality*. New York: Oxford University Press.

Synnott, Anthony. 1989. "Truth and Goodness, Mirrors and Masks – Part I: A Sociology of Beauty and the Face." *British Journal of Sociology* 40(4): 607–36.

Synnott, Anthony. 1990. "Truth and Goodness, Mirrors and Masks – Part II: A Sociology of Beauty and the Face." *British Journal of Sociology* 41(1): 55-76.

Tooby, John and Leda Cosmides. 1989. „Evolutionary Psychology and the Generation of Culture, Part I: Theoretical Considerations." *Ethology and Sociobiology* 10(1):29–49.

Townsend, John Marshall and Gary D. Levy. 1990. "Effects of Potential Partners' Physical Attractiveness and Socioeconomic Status on Sexuality and Partner Selection." *Archives of Sexual Behavior* 19(2):149–64.

Trivers, Robert L. 1971. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology* 46(1):35–57.

Trivers, Robert L. 1972. "Parental Investment and Sexual Selection." Pp. 136–179 in *Sexual Selection and the Descent of Man 1871–1971* edited by Bernard Campbell. Chicago: Aldine

Udry, Richard J. 1995. "Sociology and Biology. What Biology Do Sociologists Need to Know?" *Social Forces* 73(4):1267–78.

Webster, Murray Jr. and James E. Driskell Jr. 1983. "Beauty as Status." *American Journal of Sociology* 89(1):140–65.

Webster, Murray Jr. and Stuart J. Hysom. 1998. "Creating Status Characteristics." *American Sociological Review* 63(3):351–78.

Wilson, Rick K. and Catherine C. Eckel. 2006. "Judging a Book by its Cover: Beauty and Expectations in the Trust Game." *Political Research Quarterly* 59(2):189–202.

Wolbring, Tobias. 2010a. "Physische Attraktivität, Geschlecht und Lehrveranstaltungsevaluation. Eine Replikationsstudie zu den Befunden von Hamermesh und Parker (2005) und Klein und Rosar (2006) mit Hilfe von Individualdaten." *Zeitschrift für Evaluation* 9(1):29–48.

Wolbring, Tobias. 2010b. "Weshalb die Separierung von Produktivitätseffekten und Diskriminierung bei der studentischen Lehrveranstaltungsbewertung misslingt. Selektive Stichproben, fehlende Drittvariablenkontrolle und die Konfundierung von Effekten." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62(2): 317–26.

Wolbring, Tobias. 2012. "Class Attendance and Students' Evaluations of Teaching. Do No-Shows Bias Course Ratings and Rankings?" *Evaluation Review* 36(1):72–96.

Wolbring, Tobias and Anja Hellmann. 2010. "Attraktivität, Reziprozität und Lehrveranstaltungsevaluation. Eine experimentelle Untersuchung." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62(4):707–30.

Wolf, Naomi. 1991. *The Beauty Myth. How Images of Beauty Are Used Against Women*. New York: William Morrow and Company.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

Yurek, Leo A., Joseph Vasey and Donna S. Havens. 2008. "The Use of Self-Generated Identification Codes in Longitudinal Research." *Evaluation Review* 32:1–18.

Zebrowitz, Leslie A. 1997. *Reading Faces: Window to the Soul?* Boulder: Westview Press.

Figure 1. Directed Acyclic Graph for the Effect of Physical Attractiveness on SET

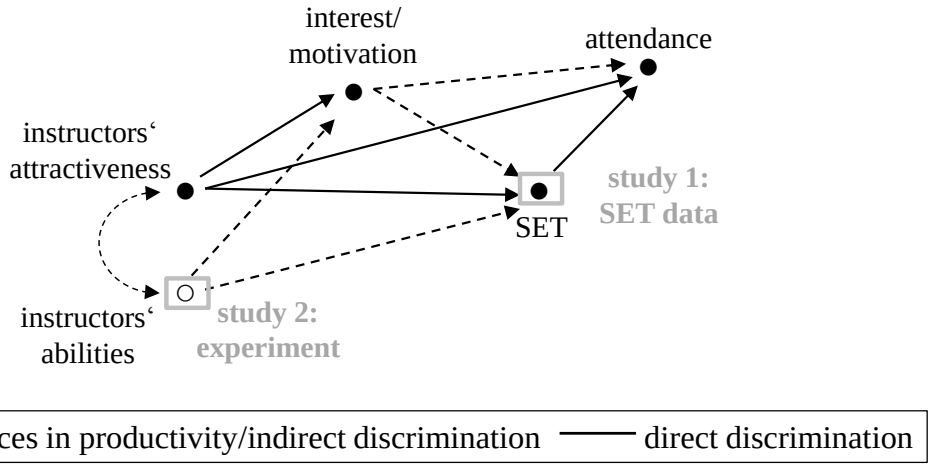
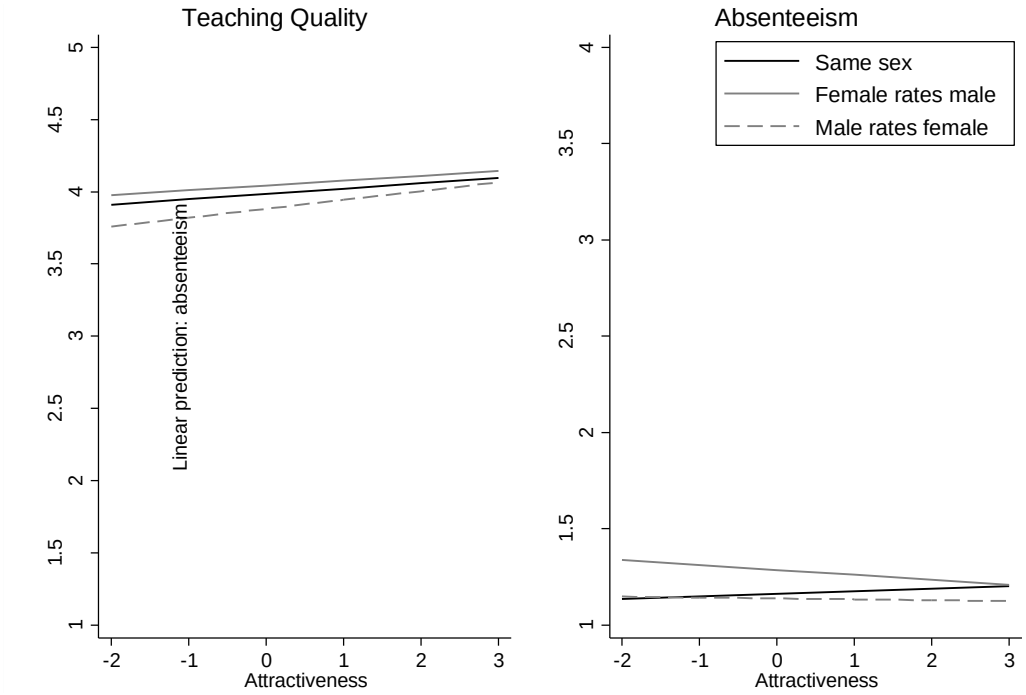


Table 1. Estimated Effects of Attractiveness and Intersexual Attraction

	‘Overall Rating: Teaching Quality’			**‘Number of Classes Skipped’**			
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Physical attractiveness (centered around the mean)	.062*** (.009)	.037*** (.010)	.039** (.015)	-.081*** (.015)	-.070*** (.015)	-.060*** (.016)	-.037 (.025)
Overall rating: teaching quality					-.195*** (.027)	-.145*** (.028)	-.146*** (.028)
Female instructor		-.068** (.023)	-.070** (.023)			-.158*** (.037)	-.175** (.039)
Interactions: Attractiveness							
*female student & male instructor			-.005 (.019)				-.045 (.033)
*male student & female instructor			.006 (.040)				-.032 (.058)
Further controls added	no	yes	yes	no	no	yes	yes
Constant	4.006*** (.00003)	3.569*** (.108)	3.569*** (.108)	1.224*** (.00003)	2.004*** (.111)	2.069*** (.236)	2.067*** (.236)
N_i	10,208	10,208	10,208	10,159	10,159	10,159	10,159
N_j	5,439	5,439	5,439	5,419	5,419	5,419	5,419
R^2_{within}	.009	.221	.221	.006	.019	.107	.107
Rho	.461	.498	.497	.533	.534	.570	.570

Note: Linear regression models with student fixed effects and robust standard errors. Nonstandardized coefficients. Standard errors in parentheses. Models 2, 3, 6, and 7 contain the following controls: performance record, department, course day and time, summer term, course pace, course difficulty, prior interest, preparation for the course, courseload, workload, and semester of study (for details see Table A1). $^{+}$ p <.1. *p < .05. **p < .01; ***p < .001 (two-tailed tests).

Figure 2. Intersexual Attraction Effects on Course Ratings and Absenteeism



Note: Ratings of teaching quality and absenteeism were predicted based on simple linear models similar to models 3 and 7 in table 1.

Table 2. Estimated Effects of Attractiveness and Blatant Misconduct

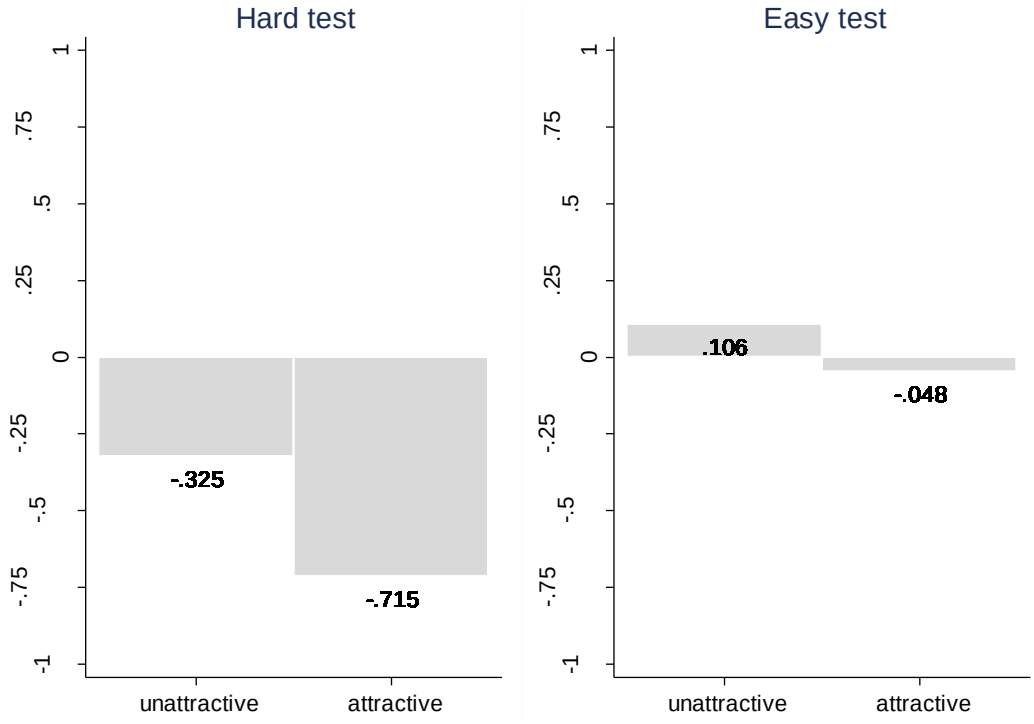
	'Overall Rating: Teaching Quality'			
	Model 1	**Model 2**	**Model 3**	**Model 4**
Physical attractiveness (centered around the mean)	0.035*** (.010)	0.037*** (.010)	0.035*** (.010)	0.035*** (.010)
Blatant Misconduct				
Course too fast (1=yes)	-0.360*** (0.075)	-0.350*** (0.073)	-0.360*** (0.075)	-0.360*** (0.075)
Course too difficult (1=yes)	-0.436** (0.140)	-0.444** (0.139)	-0.435** (0.141)	-0.437** (0.140)
Inadequate answers(1=yes)	-0.938*** (0.118)	-0.942*** (0.118)	-0.938*** (0.118)	-0.964*** (0.154)
Interactions: Attractiveness				
*course too_fast (1=yes)		-0.079 (0.056)		
*course too difficult (1=yes)			-0.0002 (0.118)	
*inadequate_answers(1=yes)				-0.064 (0.175)
Further controls added	yes	yes	yes	yes
Constant	4.340*** (0.089)	4.339*** (0.089)	4.340*** (0.089)	4.340*** (0.089)
N_i	10,155	10,155	10,155	10,155
N_j	5,415	5,415	5,415	5,415
R^2_{within}	.223	.223	.223	.223
Rho	.499	.498	.499	.498

Note: Linear regression models with student fixed effects and robust standard errors. Nonstandardized coefficients. Standard errors in parentheses. All models contain the following controls: performance record, department, course day and time, summer term, prior interest, preparation for the course, courseload, workload, and semester of study (for details see Table A1). $^{+}$ p <.1. *p < .05. **p < .01; ***p < .001 (two-tailed tests).

Figure 3. Experimental Stimuli: Photos of the Lecturers

Note: The photographs were taken from the site http://www.beautycheck.de.

Figure 4. Beauty Effects and Test Difficulty



Note: Difference of ratings is defined as rating(test prior to SET) – rating(test after SET). Positive (negative) values indicate better (worse) ratings if the SET takes place after the test.

Table 3. Estimated Effects of Attractiveness and Interaction Effects with the Timing and Difficulty of a Test

	SET before test				SET after test			
			Male rates female	Female rates male			Male rates female	Female rates male
'Overall rating: lecture quality'	**Model 1**	**Model 2**	**Model 2a**	**Model 2b**	**Model 3**	**Model 4**	**Model 4a**	**Model 4b**
Photograph (ref.: unattractive female)								
Attractive female	0.391+	0.395+	0.374		-0.036	0.279	0.311	
	(0.218)	(0.222)	(0.323)		(0.187)	(0.198)	(0.298)	
Attractive male	0.152	0.146		0.158	-0.391*	0.033		0.151
	(0.209)	(0.212)		(0.271)	(0.189)	(0.189)		(0.249)
Unattractive male	0.050	0.043			-0.206	-0.060		
	(0.209)	(0.210)			(0.189)	(0.145)		
# of correct answers (easy test)		0.007	-0.028	-0.050	0.158***	0.067**	0.040	0.020
		(0.030)	(0.066)	(0.050)	(0.027)	(0.021)	(0.049)	(0.036)
# of correct answers (hard test)		-0.033	-0.207	0.005	0.257***	0.124*	-0.072	0.201*
		(0.092)	(0.183)	(0.154)	(0.074)	(0.060)	(0.136)	(0.096)
Timing of the test (1= test prior to SET)						-0.067	0.221	-0.399
						(0.144)	(0.291)	(0.260)
Interactions: attractive female								
*hard test prior to SET						-0.509+	-0.704	
						(0.299)	(0.504)	
*easy test prior to SET						-0.008	-0.273	
Interactions: attractive male						(0.290)	(0.463)	
*hard test prior to SET						-0.549+		-0.546
						(0.295)		(0.405)
*easy test prior to SET						-0.201		0.256
						(0.284)		(0.414)
Constant	3.015***	3.017***	3.232***	3.287***	2.284***	2.735***	2.878***	2.907***
	(0.150)	(0.225)	(0.415)	(0.327)	(0.200)	(0.165)	(0.325)	(0.254)
N	162	162	36	57	161	323	71	111
R^2	.023	.027	.083	.042	.206	.095	.095	.115
Adjusted R^2	.005	-.004	-.003	-.012	.180	.066	.066	.064

Note: Linear regression. Nonstandardized coefficients. Standard errors in parentheses. $^{+}$ p <.1. *p < .05. **p < .01; ***p < .001 (two-tailed tests).