

# Olympic Medals as Fruits of Comparison? Assimilation and Contrast in Sequential Performance Judgments

Lysann Damisch and Thomas Mussweiler  
University of Cologne

Henning Plessner  
University of Heidelberg

The authors investigated the evaluative consequences of sequential performance judgments. Recent social comparison research has suggested that performance judgments may be influenced by judgments about a preceding performance. Specifically, performance judgments may be assimilated to judgments of the preceding performance if judges focus on similarities between the two. If judges focus on differences, however, contrast may ensue. The authors examined sequential performance judgments, using data gathered from the 2004 Olympic Games as well as data gathered in the laboratory with students or experienced gymnastics judges as participants. Sequential performance judgments were influenced by previously judged performances, and the direction of this influence depended on the degree of perceived similarity between the successive performances.

**Keywords:** sequential performance judgments, comparison processes, judgments in sport, assimilation, contrast

Whether people's performances lead to success or failure often depends on how others evaluate and judge these performances. Oftentimes, such performance judgments are made in sequence, so that multiple performances are judged in successive order. In the case of a job interview, for example, personnel managers welcome several candidates at the interview—one after another. Successively, they form an impression of each of the applicants in order to make a final decision about whom to hire. One realm that is particularly rich with examples of sequential performance judgments is the realm of sport. In many sports, an athlete's performance cannot be measured in terms of speed, distance, or points, as is the case in racing, swimming, or tennis, for example. In sports like figure skating, springboard diving, or gymnastics, however, the performances of consecutively starting athletes have to be assessed by judges (Landers, 1970). In fact, according to Stefani (1998), almost one third of all sports that are recognized by the International Olympic Committee have a performance-rating system in which judging plays a major role. Thus, in many sports, success and failure are determined by sequential performance judgments that are made by experienced judges.

How does this sequential nature of performance judgments—in sports and otherwise—influence the outcome of the judgment

process? Whether judgments of preceding stimuli in a sequence influence the rating of a target stimulus has been the focus of much empirical research, mostly in the field of psychophysics (Eiser, 1990; Wedell & Parducci, 2000). This research has repeatedly shown that judgments are affected by the context in which the target stimulus is presented, and thus by the nature of the stimuli that are encountered and judged prior to the target. In this literature, contrast effects have consistently been found (see Wyer & Srull, 1989). In particular, judgments of moderate stimuli in a series are often displaced away from extreme or distant stimuli. This effect has been shown across different physical dimensions, such as weight (D. R. Brown, 1953; Sherif, Taub, & Hovland, 1958), lengths of lines (Krantz & Campbell, 1961), sizes of squares (Parducci, Perrett, & Marsh, 1969), and shape (Helson & Kozaki, 1968). In much the same way, judgments about target properties that are less psychophysically and more socially defined are often contrasted away from previously encountered stimuli. In fact, such contrast effects have been observed for judgments about attitudes (Hovland, Harvey, & Sherif, 1957), the pleasantness of facial expressions (Manis, 1971), happiness (Parducci, 1968; Wedell & Parducci, 1988), attractiveness (Kenrick & Gutierrez, 1980; Kernis & Wheeler, 1981; Wedell, Parducci, & Geiselman, 1987), and fairness (Mellers, 1986).

Many of these contrast effects are in line with one prominent theoretical perspective, namely Parducci's (1963, 1965) range-frequency model, which assumes that target ratings are jointly determined by the relationship of the target stimulus to the two endpoints of the stimulus range and the relative frequency of the presented contextual stimuli. Sequential judgments, however, do not inevitably lead to contrast. At least in some studies, the opposite judgmental consequence, namely assimilation, has also been reported, so that under specific conditions, judgments about the target stimulus are moved toward a previously judged stimulus. This has been demonstrated, for example, with judgments about weight (Parducci & Marshall, 1962), loudness (M. C. King &

---

Lysann Damisch and Thomas Mussweiler, Institute for Psychology, University of Cologne, Cologne, Germany; Henning Plessner, Department of Social Psychology, University of Heidelberg, Heidelberg, Germany.

The present research was supported by a grant from the German Science Foundation (Mu1500/2). We thank the members of Social Cognition Cologne for stimulating discussions of this work. We are particularly indebted to Christina Anderson, Johannes Barthel, Darja Carl, Verena Oberst, Michael Schmitt, Maren Strauch, Julia Volkert, and Mario Zoll for their help in data collection.

Correspondence concerning this article should be addressed to Lysann Damisch, Institut für Psychologie, Erziehungswissenschaftliche Fakultät, Universität zu Köln, Gronewaldstraße 2 50 931, Köln, Germany. E-mail: lysann.damisch@uni-koeln.de

Lockhead, 1981; Ward & Lockhead, 1971), and line length (Ward & Lockhead, 1971). Assimilation also occurs for social stimuli, such as in ratings of attractiveness (Geiselman, Haight, & Kimata, 1984; Sigall & Landy, 1973). Such assimilation effects are more difficult to explain from a range-frequency perspective. Even though there have been different suggestions to theoretically explain assimilation effects in sequential judgments (e.g., Anderson, 1981; D. L. King, 1988), no approach applied to sequential judgments has yet been able to combine the diverse findings of assimilation and contrast.

In this article, we attempt to remedy this shortcoming by applying a theoretical framework developed in the realm of social comparison research to explain assimilative as well as contrastive outcomes in sequential judgment making. Therefore, we made specific assumptions about the underlying processes involved in sequential judgment making. Drawing from these process assumptions, we predicted and tested the specific judgmental outcomes and investigated which moderating factors determine their direction.

### Selective Accessibility in Sequential Judgments

Sequential judgments have a unique character in that they involve successive evaluations of several targets on one dimension. With the exception of the first judgment, for every judgment of a sequence it is true that the judge has already evaluated another target on that same dimension just a short time ago. The knowledge the judge has activated to make that previous judgment is highly accessible at the time the target judgment has to be made. As a consequence, this knowledge is likely to influence the subsequent target judgment. This is in line with one of the core principles of social cognition research, namely that accessible knowledge is especially likely to influence a judgment at hand (Higgins, 1996; Higgins, Rholes, & Jones, 1977). Thus, the evaluation of a target at almost any point of the sequence is likely to be affected by the information that was activated during the preceding judgment of another target on that dimension.

In what way can the information gathered from previously encountered and judged stimuli influence the target judgment? One plausible mechanism could be an underlying comparison process between prior stimuli and the target. The highly accessible previously encountered objects or persons are likely to serve as comparison standards against which the target is compared.

How do such comparisons influence target judgments? At first sight, the evidence that has accumulated in research on psychophysical judgment (M. C. King & Lockhead, 1981; Parducci et al., 1969; Parducci & Wedell, 1986; Ward & Lockhead, 1971), social comparison (Brewer & Weber, 1994; J. D. Brown, Novick, Lord, & Richards, 1992; Buunk, Collins, Taylor, VanYperen, & Dakof, 1990; Lockwood & Kunda, 1997), and person perception (Herr, 1986; Herr, Sherman, & Fazio, 1983; Higgins & Lurie, 1983) suggests that the direction of the comparison influence on the target judgment may be hard to predict. As is true for psychophysical judgments, research on social comparison and person perception demonstrates that comparisons may lead to assimilation as well as contrast. For example, the same target person is judged to be more hostile in the context of rather hostile others (e.g., Joe Frazier) than in the context of rather peaceful others (e.g., Robin Hood; Herr, 1986), and the self is judged to be more athletic in

comparison with an athletic person than in comparison with an unathletic standard (Mussweiler & Strack, 2000). In other studies, however, comparisons yield the opposite outcome by producing contrast effects in target evaluations. In a study by Pepitone and DiNubile (1976), for example, participants rated the same crime as more severe after an initial mild crime than after a truly cruel crime. In another study, participants judged themselves to be less competent in the context of a competent person than in comparison with an incompetent standard (Morse & Gergen, 1970).

To integrate these diverse findings, Mussweiler (2003) proposed a selective accessibility model of comparison consequences. From this theoretical perspective, the main moderating factor that determines assimilative versus contrastive comparison consequences is the perceived similarity between the standard and the target. Depending on the overall perceived similarity between a given target and a particular standard, judges engage in one of two alternative comparison processes. If similarities between target and standard stand in the foreground, then judges engage in the process of similarity testing by selectively searching for information that indicates that target and standard are indeed similar. Thus, target knowledge consistent with the standard becomes accessible. If differences between target and standard stand in the foreground, however, judges engage in the process of dissimilarity testing by selectively searching for information to indicate that the target and standard are different. Thus, accessibility of standard-inconsistent target knowledge increases. The information gathered during the process of hypothesis testing is then used to make the target judgment, as it is highly accessible at this point in time. As a consequence, similarity testing leads to assimilation, whereas dissimilarity testing leads to contrast (for a more detailed description of the model, see Mussweiler, 2003).

In the realm of social comparison research, manifold empirical support for the basic assumptions of the selective accessibility model has already been gathered. Thus, it has been shown that perceived similarity between the standard and the self influences the judgment of the self (Mussweiler, 2001a; Mussweiler, & Bodenhausen, 2002). Mussweiler (2001a), for example, used a procedural priming task to experimentally manipulate whether participants primarily focus on similarities or on dissimilarities during a comparison. The result of a subsequent social comparison task showed that participants assimilated their self-judgments to a given standard if they were made to focus on similarities. Participants contrasted their self-judgments away from the comparison standard, however, if they had been manipulated to focus on dissimilarities. Thus, an induced focus on similarities versus differences determines whether assimilative versus contrastive comparisons result. In turn, assimilative versus contrastive comparisons are accompanied by a generalized focus on similarities versus differences. More specifically, after engaging in an assimilative social comparison, participants focused on similarities in a subsequent unrelated picture comparison task. After engaging in a contrastive social comparison, however, participants focused on differences in a picture comparison (Mussweiler, Rüter, & Epstude, 2004a). Taken together with a large body of additional evidence (for a recent review, see Mussweiler, *in press*), these social comparison findings suggest that a focus on similarities versus differences critically determines whether a comparison leads to assimilation or contrast.

On the basis of these theoretical considerations and empirical findings, we assumed that whether sequential judgments lead to assimilation or contrast equally depends on whether judges focus on similarities or differences between the target and previously encountered stimuli.

### Assimilation and Contrast in Sports Performance Judgments

Out of the manifold examples for the occurrence of sequential judgments in daily life, we chose to investigate this possibility in the realm of sports. Examining theoretical questions about consequences and moderators of sequential judgments in an applied field like sports has a notable advantage. In particular, it allowed us to examine the nature of sequential judgments in a real-world laboratory. The results we received, therefore, possess a high ecological validity, because the judgments were reflective of those encountered in a natural setting. In this way, we ruled out possible doubts that our findings, as well as findings from previous research on the selective accessibility model exclusively investigated in the controlled and somewhat artificial context of the laboratory, actually are applicable to real-world phenomena (Gilovich, 1984).

Judgments in sports follow the same general principles as social judgments in other areas (Plessner, 2005, Plessner & Haar, *in press*). Under suboptimal conditions, judgments are often influenced by normatively irrelevant factors (Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980). Often performance judgments in sports have to be made under suboptimal conditions. Judgments in the field of gymnastics are one case in point. Gymnastics judges have to accomplish their task despite long-lasting competitions, the high speed of performances, and the complex and cognitively demanding regulations for judging (Salmela, 1978a, 1978b). As a consequence, performance judgments are often influenced by normatively irrelevant factors, such as feedback about other judges' scores leading to conformity (Scheer, Ansorge, & Howard, 1983), influences of order-related expectancies (Plessner, 1999), and effects of prior processing (Ste-Marie & Lee, 1991; Ste-Marie, Valiquette, & Taylor, 2001). As is true for other applied fields (e.g., Englich & Mussweiler, 2001), these influences occur not only for lay people but also for experts (Plessner & Schallies, 2005; Scheer, 1973; Ste-Marie & Lee, 1991; Wilson, 1976). This suggests that potential consequences of sequential judgments are also likely to substantiate in performance judgments in sports.

Imagine a judge in the field of gymnastics who has to assess the performances of many athletes for several hours, all starting one after another with a delay of just a few seconds. The official rules for judging are very difficult (O'Brien, 1991; Salmela, 1978a, 1978b). The speed at which athletes perform their routines is so high that it is basically impossible for the judge to detect every necessary detail (Salmela, 1978a, 1978b). Still, the judge does not have much time to announce the final score. In such a situation, we suggest, the judge uses information about the previously evaluated athlete to make a target judgment by comparing the two gymnasts with each other.

Applying the theoretical perspectives of the selective accessibility model to performance judgments in sports suggests that whether sequential judgments lead to assimilation or contrast depends on whether similarities or dissimilarities between two

consecutive athletes stand in the foreground. Let's imagine the situation of judges in gymnastics, for example. To the extent that they use the performance of a previous athlete as a comparison standard to judge another target athlete's performance, the judgmental outcome depends on whether both athletes are perceived as generally similar or dissimilar. Therefore, basic features of the gymnasts—age, regional provenance, general level of performance, or even such trivial emerging facts as the leotard color—may be taken into account. If judges perceive both athletes as similar, so the selective accessibility model suggests, they focus on further similarities between the two athletes. Thus, judges may look for information attesting a good performance of the target athlete when the performance of the standard athlete was at a high level. They could, for example, notice that the target athlete and the standard kept perfect body tension throughout their exercises. If the performance of the standard athlete was rather poor, judges would search for information indicating that the target athlete's performance was on a low level as well. If judges perceive the athletes as dissimilar, however, they focus on further dissimilarities between the two gymnasts. Thus, they could notice that the landing of the target athlete after a vault was a bit unsteady while the standard athlete showed a perfect landing. Conceivably, the judges' performance evaluations are influenced by the target information they sought. As a consequence, the judgment for the target athlete assimilates to the standard athlete's judgment in the case of similarity testing: The same performance of a gymnast is rewarded by more points if the performance of a previous gymnast was on a high rather than a low level. The judgment for the target athlete is contrasted away from the standard, however, if dissimilarity testing was engaged: The same performance of a gymnast is rewarded by fewer points if the performance of a previous gymnast was on a high rather than a low level.

In the present research, we did not aim to verify the specific assumptions of the selective accessibility model concerning mechanisms and consequences of comparison processes. In light of the rich amount of empirical evidence supporting this theoretical perspective (Mussweiler, 2003, *in press*), our goal was to apply the described assumptions of the model to a new kind of judgment, namely to sequential performance judgments. Thus, instead of testing the specific process assumptions of the model, we used the selective accessibility model to derive hypotheses about assimilation and contrast effects in sequential performance judgments in sport.

### The Present Research

In the present research, we examined whether sequential judgments in sports are influenced by performances of preceding athletes, and if so, how they are influenced. In Study 1, a first exploratory study, we analyzed the genuine gymnastics judgments of the 2004 Athens Olympic Games to detect correlations between the single scores. In Study 2, we used an experimental design to test our hypothesis that sequential athletic performance judgments are affected by previously judged athletes. Furthermore, our research examined the direction in which a judgment is influenced by preceding targets. Evidence from the social comparison literature shows that an important factor moderating the judgmental outcome of a comparison is the degree of perceived similarity between the target and the standard (see Mussweiler, 2003). In



addition, these findings suggest that perceived similarity is often the default. It has been shown that in many situations, judges initially focus on fundamental ways in which the target and the standard of a comparison are similar (e.g., Gentner & Markman, 1994, 1997). This may be because it might initially be necessary to establish a structural alignment between the target and the standard to compare them with each other. On the basis of that alignment, the focus of the comparison is defined and guides judges' attention to the critical features (Gentner & Markman, 1997; Medin, Goldstone, & Gentner, 1993). Thus, under most circumstances, judges are led toward similarity testing (Mussweiler, 2003). This assumption is further supported by evidence in social comparison research suggesting an initial focus on similarities in the comparison process (Lockwood & Kunda, 1997; Nosanchuk & Erickson, 1985). The tendency to initially focus on fundamental ways in which the target and standard are similar is further strengthened by the fact that judges typically select standards that are similar to the target (Festinger, 1954; Smith & Zárate, 1992; Suls, Gastorf, & Lawhon, 1978). On the basis of these results, we expected similarity testing to be the default mechanism and assimilation to be the default consequence in sequential judgments.

Under specific conditions, however, two consecutive targets may be perceived as dissimilar, so that dissimilarity testing and consequently contrast ensues. We manipulated the perceived similarity in two distinct ways. In Study 3, we used a procedural priming method (Mussweiler, 2001b) to induce participants to engage in similarity testing or dissimilarity testing. In Study 4, we manipulated a salient feature of the two athletes to be judged by varying their alleged nationalities. In both studies, we expected to find assimilative target evaluations when the standard and the target were perceived as similar and contrast when they were perceived as dissimilar.

With this research, we attempted to advance previous work in at least three important ways. First, to date, research on sequential judgments has mostly been conducted with unidimensional and relatively simple physical stimuli, such as target weights (D. R. Brown, 1953). Judgments about such psychophysical objects are more directly based on sensory input than judgments about more complex entities, such as performance judgments in sports (Strack, 1992). The present research thus extends previous research on sequential judgments by focusing on complex, multidimensional, and dynamic targets.

Another important difference between previous research on sequential judgments in psychophysics and the present research has to do with the format in which the judgmental responses are assessed. Most of the accumulated evidence in research on social judgments has been gathered on the basis of subjective response scales (e.g. Kenrick & Gutierrez, 1980; Parducci & Wedell, 1986; Wedell et al. 1987). The use of subjective scale responses to report a judgment, however, involves communicational influences, which potentially alter the judgment itself (Ostrom & Upshaw, 1968; Parducci, 1965; Upshaw, 1969; Wyer & Srull, 1989). One influence, for example, that changes judgmental response effects without reflecting actual changes in stimulus representation is the anchoring of a given response scale by a reference point (Ostrom & Upshaw, 1968). In sequentially judging the weight of a series of stimuli, an extreme context weight could be used to anchor the given response scale in its direction, thus moving the target judgment to the opposite end of the scale and producing a contrast

effect (Biernat, Manis, & Nelson, 1991; Ostrom & Upshaw, 1968). In the present research, performance judgments for athletic performances were not assessed with subjective judgments (e.g., How well did the athlete perform? 1 = *not well at all*, 9 = *very well*), but rather with objective judgments (e.g., How many meters did the athlete jump?). Such objective judgments are unlikely to be influenced by referent point use because they are externally anchored by consensual standards and thus give no room for interpretations (Biernat et al., 1991). As a consequence, judgmental effects that are obtained on such objective judgments more readily reflect actual changes in the evaluation of the target.

Finally, the present research also attempts to advance research on selective accessibility mechanisms in an important way. It is interesting to note that, to date, the development and verification of the selective accessibility model has always taken place in the realm of social comparison research. Even though it has been claimed that the selective accessibility model is applicable for all kinds of comparative judgments, it has not directly been tested outside of a social comparison research paradigm. In social comparison, defined as a comparison of the self with another standard in order to make a self-judgment (Festinger, 1954), motivational influences such as self-enhancement or self-protection are likely to be involved (Collins, 1996; Wheeler, 1966; Wills, 1981). This is not the case for judgments about other people's performances, which are examined in the present research. Therefore, the present research has the potential to substantiate the claim that selective accessibility mechanisms are not limited to social comparisons.

## Study 1

If judgments about a particular performance are indeed systematically influenced by the preceding performance, then this should be apparent in a correlation of consecutive performance judgments. More specifically, to the extent that similarity testing constitutes the default mechanism underlying such sequential judgments, as suggested by the findings mentioned earlier, a positive correlation between consecutive performances should ensue: The better the preceding athlete's performance, the better the target athlete's performance should be judged. In Study 1, we tested this assumption by analyzing the performance judgments of the gymnastics competitions of the 2004 Olympic Games in Athens, Greece. Given the specific situational conditions in gymnastics competitions, we particularly expected an assimilative influence based on similarity testing to be the default outcome for these judgments. Here a natural focus on similarities seemed to be especially likely, as competitors naturally shared a host of features. More specifically, participants within one event of the competition are of the same gender, belong approximately to the same age group, present similar parts of a routine at the same apparatus, and perform at a similar high level of quality. Even their competition clothing is highly similar because the international rules of gymnastics judging (Fédération Internationale de Gymnastique [FIG], 2001) include requirements for the correct clothing for each apparatus. Even though single features (e.g., their nationalities) differ among the athletes, in this natural situation they are unlikely to be sufficiently salient to overcome the strong tendency to initially focus on similarities.

## Method

For a first exploration, we selected all final scores of all athletes in the gymnastics events of the 2004 Olympic Games in their effective starting order ("Athens 2004," n.d.). This included athletes' scores in the preliminary rounds, the team finals, the individual all-around finals, and all apparatus finals for both male and female gymnasts. Taken together, we collected 1,373 Olympic gymnastics judgments.

We hypothesized that sequential judgments, such as the gymnastics judgments in the Olympics, are influenced by the performances of preceding gymnasts in an assimilative way. To examine this possibility, we performed several Pearson product-moment correlations, using the final gymnastics judgments of the athletes and the scores of their respective predecessors as dependent variables. To get a sense of the magnitude of that influence, we calculated  $N - 1$  (last performance just prior to an athlete) as well as  $N - 2$  (next to last performance prior to an athlete) and  $N - 3$  (last but two performances prior to an athlete) correlations. Hypothesizing that the gymnast starting right before an athlete whose performance has to be judged would regularly be used as the comparison standard, we expected the influence of a preceding performance to decrease with an increase of preceding performances. Thus, the  $N - 1$  correlation should be stronger than the  $N - 2$  and the  $N - 3$  correlations. For the different analyses, we had to exclude a certain number of judgments ( $N - 1$  correlation: 64 exclusions;  $N - 2$  correlation: 128 exclusions;  $N - 3$  correlation: 192 exclusions) because they had been the first, second, or third performances of a new event or a new subdivision, and thus no antecedent scores were available for them.

## Results and Discussion

Confirming our hypothesis, the  $N - 1$  correlational analysis including all events revealed a positive correlation between the target athletes' judgments and the judgments of the gymnasts starting right before them,  $r(1307) = .30, p < .01$ . We obtained a similar positive correlation for the  $N - 2$  correlational analysis,  $r(1243) = .26, p < .01$ , as well as for the  $N - 3$  correlation,  $r(1179) = .18, p < .01$ . Whereas the kind of relationship between prior performances and a target judgment remains the same, these results also demonstrate that the magnitude of this relation decreases with increasing numbers of athletes that have to be judged in between. Thus, whereas the difference between the  $N - 1$  and the  $N - 2$  correlations does not reach significance,  $z = 1.02, ns$ , the  $N - 3$  correlation is significantly smaller than the  $N - 1$  correlation,  $z = 3.23, p < .01$ , and the  $N - 2$  correlation,  $z = 2.19, p < .05$ . These results demonstrate a positive relation between gymnastics judgments for prior performances and a target performance. In other words, these results demonstrate that the score of a gymnast increases with increasing scores of his or her immediate predecessor and decreases with decreasing scores of his or her predecessor. Gymnasts are thus likely to receive a higher score if the score of the gymnast starting right before them was high rather than low. In this respect, the judgment for a gymnast is assimilated toward that of the preceding gymnast, even though, given the official guidelines of gymnastics judging, they should be independent (FIG, 2001). Moreover, we found that this influence was the strongest for the judgment immediately before a given judgment, indicating that this might be the one used as a comparison standard.

One alternative factor that may explain this positive correlation is the starting order of the gymnasts. In some parts of the Olympic gymnastics competitions, the starting order of the gymnasts within

their teams was determined by their coaches (e.g., preliminary rounds, team finals), who tend to place gymnasts in rank order from poorest to best for competition in each event (Plessner, 1999; Scheer, 1973; Scheer & Ansorge, 1975). In other events (e.g., apparatus finals), the starting order was predetermined by the athletes' scores in the preliminary round: The gymnast with the lowest score started first, and the gymnast with the highest score started last. Thus, in all of these competitions, judgments may depend on the starting order, in that performances were expected to improve with later starting positions. This influence would also produce a positive correlation between consecutive performances as detected by us. To examine whether this influence was responsible for the obtained correlation, we conducted another correlational analysis in which we statistically controlled for the influence of the starting order. Therefore, we conducted partial correlations between the judgments to ensure that they were unaffected by the influence of the starting order of the gymnasts. This was achieved by eliminating the influence of the starting order on the judgments for each gymnast as well as on the judgments for their predecessors. Thus, the correlations we found in these analyses were adjusted for the influence of the starting order. To avoid a confound in combining data from events where the order of competition was affected by prior performance or by strategies of the coaches and data from competitions where this does not happen, we conducted the partial correlation for the affected events only. The partial correlation ( $N - 1$ ) including the judgments of the preliminary rounds, the team final, and the apparatus finals still revealed a positive correlation between judgments for gymnastics performances and the judgments for their immediate preceding performances,  $pr(1076) = .31, p < .01$ . Moreover, this result indicates that controlling for the influence of starting order does not decrease the positive correlation in comparison with the uncorrected  $N - 1$  correlation including these three events,  $r(1077) = .31, p < .01, z = .08, ns$ . This suggests that the obtained correlation between the target athlete's performance and the performance of the predecessor was not produced by the starting order of the gymnasts.

Taking advantage of the richness of this data, we conducted another exploratory analysis to detect whether—under specific conditions—negative correlations could be found for these Olympic judgments as well. A negative correlation between gymnasts' scores and the proceeding judgments would be a hint toward the occurrence of contrast effects. A contrast effect, in turn, should occur if subsequently starting gymnasts were naturally perceived as dissimilar. Therefore, we split the data into two groups on the basis of the more-or-less only characteristic of the gymnasts we had, namely their nationalities. A first group included all the gymnasts whose preceding athletes had the same nationalities. A second group includes all the gymnasts whose preceding athletes had a different nationality. For both groups, we ran  $N - 1$  bivariate correlation analysis. For the same-nationality group, we predicted a positive correlation between successive judgments, just as reported for the overall analysis above. For the different-nationality group, one could, in principle, expect a negative correlation to occur, thus signifying a contrast effect. If the knowledge about the gymnasts' nationalities was sufficiently salient for the gymnastics judges, then a natural focus on dissimilarities may have arisen so that the comparison between two successive gymnasts led to contrast. However, given that the available data stem from an

international competition that included 44 different participating nations, it seems likely that knowledge about the different nationalities of two gymnasts may not have been sufficiently salient to induce judges to overcome the default tendency to focus on similarities and engage in the process of dissimilarity testing. In this case, we would predict a positive correlation between succeeding judgments just as for the same-nationality group.

In line with the latter argument, the results revealed a positive correlation between the judgments of successively starting gymnasts sharing the same nationality,  $r(663) = .29, p < .01$ . For successive judgments for gymnasts who differed in their nationalities, we also found a positive correlation,  $r(642) = .30, p < .01$ . A comparison between the two coefficients of correlation revealed no difference in the strength of the obtained correlations,  $z = .26, ns$ . These results suggest that for the analyzed Olympic gymnastics judgments at hand, the information about the gymnasts' nationalities was not sufficiently salient to overcome the judges' initial focus on similarities.

In summary, the reported findings are fully in line with our hypotheses. Analyzing the scores of Olympic gymnastics judges, who can be considered highly qualified in judging because they have to meet several strict requirements of the FIG to be assigned for an Olympic competition (FIG, 2005), we found that judgments of consecutively starting athletes were positively correlated. This influence is stronger the closer in time the prior judgment is made to the target judgment. Although statistically controlling for the influence of starting order, these results show that the likelihood for an Olympic gymnast to receive a high score for his or her exercise is higher if the preceding gymnast presented a good rather than a flawed performance. Putting this positive correlation into other words, one can say that the performance judgments of consecutively starting gymnasts are linked with each other in an assimilative way. In the present data, we found no evidence for negative correlations, and thus no hint for contrast effects, even if successively starting gymnasts varied in one characteristic, their nationalities, from each other.

## Study 2

The results of Study 1 provide initial support for the occurrence of an assimilative influence of preceding performances on consecutive scores in sequential judgments. Because these data are correlational in nature, however, they do not allow for causal interpretations. To remedy this shortcoming, we conducted Study 2. In an experimental design, we tested our hypothesis that the sequential evaluation of athletic performances is influenced by the achievement of previously encountered athletes. If that were the case, then the performance of the same target should be evaluated differently depending on the standard athlete's level of performance. To test our hypothesis, we asked participants to evaluate the athletic performance of a pictured target athlete after they had judged the performance of a high versus a low standard along the same judgmental dimension. In light of the results of Study 1 and the presented theoretical remarks suggesting the process of similarity testing to be the default process, we expected an assimilation effect to be the outcome of this study. In other words, the evaluation of a well performing preceding athlete should lead to higher judgments of the subsequent target athlete than the evaluation of a poor performing preceding athlete.

## Method

**Participants.** We recruited 42 male students at the University of Würzburg as participants and randomly assigned them to one of two experimental conditions. Participants were offered a chocolate bar as compensation.

**Materials.** The materials for the judgment task consisted of three photographs, each showing a female athlete performing in the event of the long jump. All pictures were monochrome pictures, which made it impossible to tell the athletes' affiliation to a certain country or sports club. The pictured athletes were kept anonymous. None of our participants reported recognition of any of the pictured athletes. The three pictures differed with respect to the performance level of the pictured athletes. One athlete (high standard) was clearly recognizable as performing at a high level, evidently jumping quite a long distance. Another athlete (low standard) was obviously performing at a low level by doing a rather short jump. The third picture showed an athlete (target) performing at a moderate level doing a moderate jump. These photos were pretested by asking a different set of participants ( $N = 67$ ) to estimate for one picture each the achieved length of the pictured long jumps. Our analysis revealed that estimates for the high standard were higher ( $M = 5.93$  m,  $SD = 1.31$  m) than estimates for the low standard ( $M = 4.44$  m,  $SD = 0.70$  m),  $t(46) = 4.97, p < .01, r = .59$ . The mean estimates for the moderate target ( $M = 5.00$  m,  $SD = 0.91$  m) were lower than estimates for the high standard,  $t(40) = 2.60, p < .05, r = .38$ , and higher than estimates for the low standard,  $t(42) = 2.33, p < .05, r = .34$ .

**Procedure.** Participants were recruited at the university's cafeteria, asked to participate in a brief psychological test about judgments in sports, and led to a separate room where they were seated at a table. The study was performed with groups of up to 6 participants simultaneously. By reading the instruction, which was given to them in a folder, participants were informed that the purpose of the test was to investigate people's ability to judge performances of athletes on the basis of simple pictures. Therefore, it would be their task to look at pictures of different athletes and estimate the performance of each athlete carefully even if the task seemed to be very difficult. The next page contained of a picture of the first long jumping athlete, which served as the standard. Half of the participants were presented the high standard. The other half of the participants were presented the low standard. On the same page, below the picture, participants were asked to write down their estimated lengths of the pictured jump. Subsequently, all participants had to judge the jump of the moderate target athlete which was presented on the next page. The study was finished after completion of this task; thus, two judgments were required of each participant. They were then debriefed, thanked, and offered their candy.

## Results and Discussion

We expected judgments about the performance of the target athlete to be assimilated toward the performance of the standard athlete. If this was indeed the case, then the target jump should have been estimated to be longer after judging the high standard rather than the low standard. Consistent with this hypothesis, participants' target judgments yielded an assimilation effect. Specifically, participants tended to estimate the length of the target jump as further after judging the high standard ( $M = 598.33$  cm,  $SD = 127.24$  cm) rather than the low standard ( $M = 467.86$  cm,  $SD = 115.44$  cm),  $t(40) = 3.48, p < .01, r = .48$ .

The obtained effect demonstrates that sequential judgments in sports are indeed influenced by previously encountered athletic performances. Furthermore, these findings are also consistent with our hypothesis concerning the direction of that influence on the target judgment. Consistent with the correlative evidence gathered in Study 1, we find an assimilation effect to be the default effect in sequential performance judgments.



### Study 3

As demonstrated in Studies 1 and 2 assimilation often appears to be the default judgmental outcome. Nevertheless, comparisons may also lead to contrast if the characteristics of the initial situation lead judges to focus on dissimilarities between the standard and the target and, thus, to engage in the process of dissimilarity testing. We conducted Study 3 to test our assumption that the outcome of sequential performance judgments depends on whether judges focus on similarities or dissimilarities between the consecutively presented athletes.

To do so, we actively manipulated our participants' comparison focus before the actual judgment task. For this purpose, we used a procedural priming task that had been applied by Mussweiler (2001b). In an unrelated task prior to the critical judgment, participants were asked to compare two sketches with each other by listing either all similarities or all dissimilarities they could identify. Subsequently, participants judged the performances of sequentially presented athletes. We expected that the way of processing information in the picture comparison task, namely searching for similarities or dissimilarities, would become proceduralized and carry over to the subsequent judgment task. Thus, searching for similarities among the sketches in the priming task was expected to induce judges to focus on similarities in the judgment task. Focusing on dissimilarities, however, was expected to lead judges to attend to dissimilarities during the judgment task. If the evaluation of a target athlete's performance indeed depends on the participant's focus on similarities versus differences, we would expect judges to assimilate their judgment to that of the standard after they had been primed to focus on similarities. Conversely, we would expect judges to contrast their judgment away from that of the standard after they had been induced to focus on differences.

### Method

**Participants.** We recruited 72 students at the University of Würzburg as participants and randomly assigned them to one of four experimental conditions. Participants were offered a chocolate bar as compensation.

**Materials.** The materials for the procedural priming task consisted of two drawings by Albrecht Dürer. The first drawing depicts an old man kneeling at the bottom of a canyon. Next to him lies a lion, and in the far distance, one can see the roof of a house. The second picture also shows a kneeling man. He appears to be in the middle of a village, surrounded by feeding pigs.

The materials for the judgment task consisted of three short video scenes showing male athletes in the event of ski jumping. Each of the three athletes could be seen during a whole jump, including the approach, the jump, and the landing. The color of the athletes' clothing was similar for all three athletes and did not identify their nationalities. The athletes were kept anonymous, and none were recognized by our participants. Similar to Study 2, the three scenes differed from each other in the general level of performance presented by the athletes. One of the videos showed an athlete clearly performing at a high level (high standard). By landing in the middle of the hill, another athlete was clearly recognizable as performing at a low level (low standard). The third scene showed an athlete performing a moderate ski jump (target). The true length of the target ski jump, which of course was not visible for the participants, was 108 m, which was 14 m longer than the short ski jump and 14 m shorter than the long ski jump. The videos were presented as MPEG (Moving Picture Experts Group) video files (resolution: 384 × 288 pixels) on a 15-in. laptop screen, using MediaLab (Jarvis, 2002) as our research software.

**Procedure.** Participants were recruited at the university's cafeteria, asked to participate in two brief psychological studies, and led into a separate room with up to 2 other participants. Participants were given verbal instructions that the first study would be a paper-and-pencil task, whereas the second study would be performed at the computer. Participants were told that both studies were absolutely unrelated to each other. As an awareness check at the end of the study revealed, 4 participants were suspicious about the potential relatedness of the two studies. They explicitly mentioned that searching for similarities or dissimilarities in the first task had influenced their performance estimates in the second task and were thus excluded from our analyses. Participants then received a folder, containing the instruction and the materials for the picture comparison task. According to the cover story, the participants believed they were about to pretest several pictures for a later memory task. They were asked to look carefully at the two pictures and compare them with each other as precisely as possible. Half of the participants were then asked to list as many similarities between the pictures as they could find, and the other half were asked to list all the differences they could find between the pictures. There was no time limit for this task.

After completion of the first part, the experimenter started the second part at the computer. Similar to Study 2, participants were informed that the purpose of the test was to investigate people's ability to judge performances of athletes on the basis of videos. Therefore, they were asked to watch some videos presenting ski jumpers and estimate the achieved length for each of the presented ski jumps, even if the task seemed to be very difficult. Participants then started the first video, which served as the standard. Half of the participants saw the high standard jump. The other half saw the low standard jump. After giving their estimates of the achieved length of that first ski jump on a separate screen, all participants saw and evaluated the moderate target ski jump.

After completing this task, participants answered a short questionnaire including some self-descriptive information, indicating their sex, age, field of study, and current number of semesters. The procedure finished with an awareness check consisting of the following two questions: "Did you notice anything unusual?" and "Did you have the feeling that the two studies you just took part in are somehow related to each other?" If participants answered either of these questions in the affirmative, they were asked to give a written explanation detailing their concerns. They were then debriefed, thanked, and offered their candy.

In summary, the described manipulations resulted in a 2 × 2 Standard (high vs. low) × Focus (similarity vs. dissimilarity) experimental design.

### Results and Discussion

Before analyzing the data, we excluded four implausible estimates of the ski jumps' lengths. One participant, for example, estimated the length of the standard ski jump to be 800 m and that of the target ski jump to be 750 m. Taking into account that 239 m is the current world record in ski jumping ("Sport 1. The Development," 2006) these estimates are clearly implausible. In addition, we excluded two estimates that deviated from the mean by more than 3 standard deviations. Both methods are commonly used to detect and handle outliers (Bargh & Chartrand, 2000). All participants excluded from the analyses for the explained reasons were evenly assigned to the four existing experimental conditions.

The observed pattern of means demonstrates that the estimation of a ski jump's length presented after a high versus a low standard depends on whether judges were procedurally primed to focus on similarities or dissimilarities. As expected, participants who focused on similarities estimated the length of the target ski jump to be relatively higher if it was presented after the high standard ( $M = 104.36$  m,  $SD = 16.7$  m) rather than the low standard ( $M = 96.00$  m,  $SD = 17.8$  m), thus producing an assimilative tendency.

For participants focusing on dissimilarities, however, this pattern reversed into a contrast effect. In this condition, the target ski jump's length was estimated to be longer if it was presented after the low standard ( $M = 102.59$  m,  $SD = 18.7$  m) rather than the high standard ( $M = 91.31$  m,  $SD = 23.9$  m). In a  $2$  (high vs. low standard)  $\times 2$  (similarity vs. dissimilarity focus) analysis of variance (ANOVA) with the length estimation for the target ski jump as the dependent variable, this pattern produced a significant interaction effect,  $F(1, 58) = 3.87$ ,  $p < .05$ ,  $r = .25$ . None of the remaining effects were significant: main effect of focus,  $F(1, 58) = .41$ , *ns*; main effect of standard,  $F(1, 58) = .08$ , *ns*.

The findings of Study 3 provide further support for our assumption that sequential performance judgments in sports are influenced by the performance of preceding athletes. Congruent with our findings in Study 2, the results demonstrate that judgment of an athlete's performance depends on the context in which it is presented. Thus, the judgmental outcome is a different one, depending on whether another well-performing or another poor-performing athlete is presented and judged prior to the target. The results also show that the performance quality of the presented standard is not the only factor determining the judgmental outcome. The direction in which sequential judgments are influenced is determined by whether judges focus on similarities or dissimilarities between successively starting athletes. A judge who concentrates on similarities between the athletes assimilates his or her target judgment to that of the previous athlete. For a judge focusing on dissimilarities, however, the judgmental outcome of the sequential performance evaluation is a contrast effect.

#### Study 4

In the previous study, we actively manipulated whether participants focused on similarities or dissimilarities. In real life situations, however, it is unlikely that judgments would be based on a direct focus manipulation of the kind used in Study 3. Rather, the direction of the judgmental outcome is determined by the degree of the perceived similarity between the target and the standard. Many examples supporting this assumption can be found in the recent literature (Brewer & Weber, 1994; Herr, 1986; Mussweiler, 2001b; Mussweiler, Rüter, & Epstude, 2004b). These examples are also helpful in determining which features influence whether two targets are perceived as generally similar or dissimilar. Prominent moderating factors include the extremity of the standard (Herr, 1986; Herr, Sherman, & Fazio, 1983; Mussweiler et al., 2004a), the ambiguity of the target (Herr et al., 1983; Pelham & Wachsmuth, 1995), or the group category membership of the target (Brewer & Weber, 1994; Mussweiler & Bodenhausen, 2002). In our final experiment, we manipulated the perceived similarity of two athletes by varying their category membership on the basis of a very salient feature.

Study 4 was conducted in the field of gymnastics. Similar to Studies 2 and 3, the participants evaluated the performances of sequentially presented athletes. To vary the perceived similarity between the presented gymnasts, we experimentally manipulated their alleged nationalities. By giving the participants the written information about the nationalities of the gymnasts, while other characteristics were not mentioned or were controlled, this feature of the gymnasts was considered to be a very salient attribute that would be taken into account to assess the general degree of

similarity between the gymnasts. We expected the two presented gymnasts to be perceived as similar if both of them apparently had the same nationality. In this case, we predicted, judges would engage in similarity testing, hence leading to assimilation effects. If the gymnasts differed in terms of their nationalities, however, dissimilarity testing would be more likely, and thus, contrast effects would occur.

#### Method

**Participants.** We recruited 47 certified and highly qualified, male, German gymnastics judges to participate in this experiment. They were recruited at five different gymnastics events throughout Germany. During longer breaks within a given event or at the end of a competition, they were asked to participate in our study. The judges' mean age was 40 years, with the youngest participant being 17 years old and the oldest being 68 years old. They had been practicing as judges for 15 years on average (minimum: 1 year; maximum: 40 years) and, thus, were all experienced with the task. On average, they were assigned to 13 gymnastics competitions per year (minimum: 0–1; maximum: 30). In addition, 43 of the 47 judges indicated that they had been active in the sport of gymnastics themselves ( $M = 18$  years; minimum: 5 years; maximum: 40 years). Most important, as it was our minimum criterion for participation, all participants held a valid license for male gymnastics judging, ranging from licenses at the federal state level ( $N = 26$ ), the national level ( $N = 12$ ), and up to internationally valid licenses ( $N = 8$ ). One participant did not specify his license. Participants were however assigned to one of four experimental conditions.

**Material.** Judges scored routines on the vault, which were presented on videotapes. The videotapes were from the 1997 World Championships of Gymnastics in Switzerland. The relatively old recordings were chosen to avoid any effects of recognition of particular athletes. The specific materials consisted of three different gymnasts performing the same vault in terms of technical requirements and degree of difficulty (the vault called the Tsukahara stretched). Original scores were neither visible nor audible to prevent biases of the judges' scores through this kind of feedback (Scheer et al., 1983). A single scene consisted of the approach run, the vault, and the landing of the gymnast in realistic speed. The leotards of the three pictured gymnasts did not differ in color, but the athletes were nevertheless clearly recognizable as different people. The competition clothing of the gymnasts did not allow identification of the athletes' team memberships. The three presented vaults differed in quality of performance. The vault in one scene was a high-quality performance, which was pre-rated by three certified and nationally qualified gymnastics judges with a final score of 8.43 points (high standard). The quality of performance of the second vault can be described as rather low (low standard). The average final score in the prestudy was 8.17 points. The score for the third selected vault was 8.23 points in the prestudy, thus representing a moderate quality of performance (target). The videos were presented as MPEG video files (resolution:  $384 \times 288$  pixels) on a 15-in. laptop screen, using MediaLab (Jarvis, 2002) research software.

**Procedure and design.** Participants were asked to take part in a short psychological study to examine the complex cognitive processes that are involved in judging gymnastics routines. After agreeing to participate, they were led into a separate room where they were placed in front of a laptop at a table. Up to three judges participated simultaneously.

Judges were first instructed that it would be their task to judge two gymnastics performances. The judges were asked to apply the official rules concerning the evaluation of male gymnasts' routines, prescribed by the code of points (FIG, 2001). Specifically, they learned that all routines presented to them on a video screen would be at the apparatus of the vault and that the routines had to be judged in terms of their technique and body position. Participants were told that they would see every routine only once. The possibility of a replay or slow motion was not given in order to keep the experimental procedure as close to a natural competition situation



as possible. Participants were then presented with the first vault. Half of the participants saw the high standard vault; the other half saw the low standard vault. Immediately after viewing the first gymnast, judges were asked to make their judgment. After writing down the calculations of the final score, the presentation of the second vault started. This time all participants saw the same target routine. Judges were again asked to report the final scores for the second vault after viewing it.

In addition to the performance-quality manipulation of the standard, we also varied the perceived similarity between the standard and the target by manipulating the alleged nationality of the gymnast performing the first vault (the standard). Half of the participants received the written information right before viewing the first routine that the performing gymnast was Australian. The remaining judges were told that the first gymnast was from Canada instead. For both groups, the nationality of the second gymnast (the target) was Australian again. This way, for one group, both of the presented gymnasts had the same nationality (Australian). Because the gymnasts did not differ in any other salient features except their level of performance quality, judges were expected to perceive the standard and the target as similar to each other in this condition. For the remaining participants, the presented gymnasts did differ in their nationalities. Even though this group actually saw the same two routines as the first group, we expected them to perceive the standard and the target as dissimilar from each other because the information about the differing nationalities was presented in a way that made it very salient.

Australia and Canada were chosen deliberately because gymnasts from both nations are internationally well acknowledged. On the other hand, it is not possible to work out a directed hypothesis predicting which country has the better gymnasts. At the gymnastics world championship in 1997, for example, the male gymnasts of these two nations finished the team competition in Positions 17 and 18.

When the judges reported their final scores for the target, they filled in a form giving information about their sex, age, and profession, how long they had been active as a gymnastics judge, their average number of appointments as a judge per year, the amount of time they had been active as a gymnast themselves, and their license. After completion of that questionnaire, participants were debriefed and thanked. The whole procedure lasted approximately 10 to 15 min.

## Results and Discussion

According to the described design, we conducted a 2 (high vs. low standard)  $\times$  2 (same vs. different nationality) ANOVA with the final target judgments as the dependent variable. The aim of Study 4 was to replicate the findings of Study 3 in a setting containing a maximum of ecological validity. Specifically, we wanted to demonstrate that the focus of similarity that we directly manipulated in Study 3 was naturally determined by the degree of generally perceived similarity between two athletes, yielding the same judgmental consequences. The manipulation of perceived similarity was achieved by varying the alleged nationality of our standard gymnast.

As the observation of the pattern of means reveals, the consequences of judging the target routine subsequent to a high versus a low standard go in different directions, depending on whether the standard and target belong to different nationalities or to the same nationality. Consistent with our hypotheses, in the same-nationality condition, judges assimilated their target judgments to that of the standard. They rewarded the target vault with a relatively higher score after evaluating the high standard ( $M = 8.23$ ,  $SD = 0.24$ ) rather than the low standard ( $M = 8.07$ ,  $SD = 0.29$ ). For the different-nationality condition, however, the reverse pattern occurred. Here, the target vault was evaluated with higher

scores after judges had previously seen a low standard ( $M = 8.35$ ,  $SD = 0.14$ ) rather than a high standard ( $M = 8.15$ ,  $SD = 0.32$ ). The result is a contrast effect.

The ANOVA revealed that this pattern produced a significant interaction effect,  $F(1, 43) = 5.03$ ,  $p < .05$ ,  $r = .32$ . None of the remaining effects were significant: main effect of standard,  $F(1, 43) = 0.07$ , *ns*; main effect of nationality,  $F(1, 43) = 1.78$ , *ns*,  $r = .19$ .

The results of Study 4 clearly replicate the findings of the previous studies. Even though all of our participants were verifiably experienced with the task of evaluating gymnastics performances, we found their judgments to be influenced by a sequence effect of reasonable size. Specifically, our findings demonstrate that the judgment for the target gymnast highly depends on the performance of the preceding gymnast.

Our results further suggest that the direction of that sequential judgment effect is determined by the degree of perceived similarity between two consecutively presented gymnasts. Whereas in Study 1 the feature of the gymnasts' nationalities was not salient and thus not sufficiently strong to make a difference in the kind of correlation between succeeding judgments, it had a crucial impact on the outcome in Study 4, where it was the only information about the gymnasts available to the judges. Thus, we demonstrated that the manipulation of a single salient feature can be enough to influence whether two successively presented athletes are perceived as similar or dissimilar. Depending on whether those two athletes share that salient characteristic or differ in it, judges either assimilate to or contrast their target scores from the preceding judgment.

## General Discussion

Taken together, these findings indicate that sequential performance judgments in sports are influenced by previously judged athletes of the same sequence. In Study 1, we found that the Olympic gymnastics judgments of consecutively performing gymnasts are not independent from one another but rather are correlated in a positive way. As we demonstrated, this correlation can be interpreted as a tendency of the judges to assimilate a gymnast's score to that of the previous gymnast. Study 2 supports this explanation by yielding an assimilation effect as the result of an experimentally controlled sequential judgment task. The results of Studies 3 and 4 extend these findings by demonstrating that sequential judgment effects can go in opposite directions. In Study 3, we successfully applied a procedural priming method to manipulate whether judges focus on similarities or dissimilarities of the presented athletes, whereas in Study 4 the degree of perceived similarity between two athletes was manipulated by varying their ostensible nationalities. In both studies, assimilation occurred when judges focused on similarities, and contrast was the result of focusing on dissimilarities.

These findings have a number of implications for performance judgments in sports specifically and for social judgments in general. Although the present research was not designed to examine the specific mechanisms that underlie the obtained sequence effects, it demonstrates that the judgmental consequences and their moderating factor are similar to those that are typical of comparison processes. As is true for comparison consequences, it appears to be the degree of perceived similarity between two consecutively

judged targets that determines whether assimilation or contrast is the judgmental outcome. In this respect, the same basic principles that determine comparison consequences also appear to apply to sequential judgments. This concordance suggests that comparison processes contribute to the obtained effects. In fact, it is true for gymnastics judgments, for example, that the official rules encourage judges to engage in comparison processes. Specifically, to detect technical and positional execution errors of a given performance, judges have to compare the performed elements to an ideal standard (FIG, 2001). A comparison of this kind, however, would not explain the pattern of results in our studies. Rather, our findings suggest that it is not a comparison between an ideal and the target performance but a direct comparison between the target performance and a prior performance of a preceding athlete within the same sequence. This finding is supported by results of a survey by Plessner (1997), where more than half of the interviewed gymnastics judges reported that they relied on direct comparisons when judging a routine. This suggests that judges use a prior performance as a standard of comparison to evaluate a target performance. The question of which one of the preceding performances is used as a standard of comparison is not clearly solved in our research design because we had participants judge only two performances in our experiments. In principle, it is conceivable that any prior performance, such as the very first performance of a sequence or the best performance, could serve as a standard of comparison. However, we suggest that it is mostly the performance right before the target that serves as the standard, as this should be the most accessible one. This assumption is supported by our findings in Study 1. In this study, we demonstrated that the positive correlation between succeeding judgments was strongest between a target performance and the immediately preceding performance. This correlation decreased with increasing numbers of intervening judgments. Although this evidence is merely correlative in nature, we interpret this finding to support our suggestion that, by serving as a comparison standard, the performance just prior to a target performance has the strongest influence on the target judgment. This should also apply to sequential judgment situations consisting of more than two performances, as would typically be the case in real gymnastics competitions, for example.

Comparison processes thus appear to contribute to sequential judgment making. We go even further and argue that judges may spontaneously engage in such a direct comparison between succeeding performances even if they are aware of that sequential influence. This assumption is strengthened by recent findings in the field of social comparison suggesting that comparisons oftentimes just happen in a relatively spontaneous, effortless, and unintentional manner. Comparisons are not always deliberate processes that are strategically engaged. Rather, they often appear to be carried out spontaneously, even if a given standard is clearly irrelevant (Gilbert, Giesler, & Morris, 1995). Comparisons even influence judgments if people are not aware of potential standards because they were presented subliminally (Mussweiler & Englich, 2005; Mussweiler, Rüter, & Epstude, 2004b). This evidence demonstrates that oftentimes comparisons simply happen. They are spontaneously engaged whenever a potential standard is accessible. In the special case of sequential performance judgments, a host of potential standards is available in the form of previously evaluated targets. In addition, these potential standards are highly accessible because they have just been evaluated a short time ago.

This suggests that in sequential judgments, target evaluations are likely to be influenced by spontaneous comparisons with prior targets. Sequential judgments may thus essentially be comparative judgments.

From this perspective, the present results also contribute to the wide field of comparison research, which is often conducted in the context of social comparison (Festinger, 1954). Here, it has recently been suggested that the psychological mechanisms that underlie comparisons and the effects they produce differ depending on whether a comparison is explicit or spontaneous in nature (Stapel & Suls, 2004). The present findings demonstrate that this is not necessarily the case. In none of our studies were participants explicitly asked to compare the presented athletes with one another. In fact, we carefully avoided using words like *comparison* or *compare* in our instructions. Still, in all of our studies we obtained a similar pattern of judgmental consequences that has been found in studies examining consequences of explicit comparisons (e.g., Mussweiler, 2001b). This suggests that more explicit and more spontaneous instances of comparative information processing may well involve the same psychological mechanisms and yield similar judgmental effects.

In addition to comparison research and social judgment research, the present findings also are relevant for the applied field of sport, particularly for research on gymnastics judgments. In all four of our studies, we investigated the effects of sequential judgments in a realistic context across different kind of sports. Concordant with our predictions, we found for all of them that evaluating athletic performances successively led to sequential influences of preceding performances on target judgments. In the sports long jumping and ski jumping used in Studies 2 and 3, however, the impact of the investigated influence would rarely have an effect on the ascertainment of the athletes' performances in a realistic competition because their achieved distances can objectively be measured. This is not the case in Studies 1 and 4 because gymnastics performances are evaluated by human judgments. In both studies, sequential performance judgments of highly experienced gymnastics judges were investigated, thus increasing the ecological validity of this research. The findings of Study 1 and Study 4 clearly demonstrate that these judgments are not free from distorting influences, even when made by experienced judges. These results are in line with previous findings in gymnastics research (Plessner & Schallies, 2005; Scheer, 1973; Ste-Marie & Lee, 1991) and in other fields (Englich & Mussweiler, 2001; Englich, Mussweiler, & Strack, 2006), thus demonstrating that judgments of experts and lay people are influenced in similar ways. Especially in Study 4, we attempted to keep the laboratory environments as close as possible to the natural setting of judges in gymnastics competitions. Thus, in addition to the fact that all of our participants were well trained in the task of judging gymnastics performances, we used original gymnastics performances as stimuli and assessed their evaluation along the official guidelines of gymnastics judging. However, it should be noted that some differences between our laboratory setting and the real world remained. For example, judges watched the exercises on a computer screen instead of on the competition floor. In relation to that divergence, other issues—like the kind or amount of the visual information assessed, the acoustic information assessed, or feelings of pressure—also are likely to have diverged from the realistic situation of gymnastics judges.

Nevertheless, we argue that the demonstrated sequential effects are of high practical relevance for judgments in gymnastics. A closer look at the findings of Study 4 reveals that the final score for the very same gymnastics performance differs depending on the performance of a preceding athlete and the perceived similarity between the sequentially judged gymnasts. The described difference, though in opposite directions, constitutes about two tenths of a point for the similarity as well as the dissimilarity condition. At first sight, this may appear negligible, but a closer look reveals that two tenths of a point can have a crucial impact in the sport of gymnastics. In the last Olympics, for example, for each of the six apparatuses on which men perform, the score of the winner and the score of the gymnast ending at the sixth position did not differ by more than two tenths of a point (FIG, 2004). In light of the present findings, paradoxically, a gymnast who won a gold medal because he happened to start after a gymnast who was similar to him and performed well might have gone home without a medal if the gymnast before him had offered a flawed performance.

In conclusion, the present research demonstrates that sequential performance judgments are not independent from one another. Rather, a preceding judgment clearly influences the actual target judgment in predictable ways. Specifically, the direction of the influence depends on the perceived similarity between the sequentially presented stimuli. If a target is perceived as generally similar to its predecessor, assimilation effects ensue. If consecutively presented stimuli are perceived as generally dissimilar, however, contrast effects in the target judgment are expected.

To demonstrate the potential impact and practical relevance of these findings again, we conclude with another example out of the world of gymnastics. Paul Hamm, an outstanding American gymnast participating at the 2004 Olympics qualified for four out of the six apparatus finals. In none of them, however, did he win the Olympic gold medal. Taking into account that in all four competitions, he missed that goal by less than one tenth of a point, we can speculate that his dream of an Olympic gold medal in one of those finals could have come true, if he had only started after the right predecessor.

## References

- Anderson, N. H. (1981). *Foundations of integration theory*. New York: Academic Press.
- Athens 2004: *The official Website of the ATHENS 2004 Olympic Games*. (n.d.). Retrieved August 16, 2005, from [http://www.athens2004.com/en/resultsGymnastics\\_Artistic.html](http://www.athens2004.com/en/resultsGymnastics_Artistic.html)
- Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 253–285). New York: Cambridge University Press.
- Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, 60, 485–499.
- Brewer, M. B., & Weber, J. G. (1994). Self-evaluation effects of interpersonal versus intergroup social comparison. *Journal of Personality and Social Psychology*, 66, 268–275.
- Brown, D. R. (1953). Stimulus-similarity and the anchoring of subjective scales. *American Journal of Psychology*, 66, 199–214.
- Brown, J. D., Novick, N. J., Lord, K. A., & Richards, J. M. (1992). When Gulliver travels: Social context, psychological closeness, and self-appraisals. *Journal of Personality and Social Psychology*, 62, 717–727.
- Buunk, B. P., Collins, R. L., Taylor, S. E., VanYperen, N. W., & Dakof, G. A. (1990). The affective consequences of social comparison: Either direction has its ups and downs. *Journal of Personality and Social Psychology*, 59, 1238–1249.
- Collins, R. L. (1996). For better or worse: The impact of upward social comparison on self-evaluations. *Psychological Bulletin*, 119, 51–69.
- Eiser, J. R., (1990). *Social judgment*. Pacific Grove, CA: Brooks-Cole.
- Englich, B., & Mussweiler, T. (2001). Sentencing under uncertainty: Anchoring effects in the courtroom. *Journal of Applied Social Psychology*, 31, 1535–1551.
- Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin*, 32, 188–200.
- Fédération Internationale de Gymnastique. (Hrsg.). (2001). *Code de pointage—Gymnastique artistique masculine* [Code of points—male gymnastics]. Moutier, Switzerland: Author.
- Fédération Internationale de Gymnastique. (2004). *Athens Olympic Games—Men's artistic results*. Retrieved August 16, 2005, from <http://www.fig-gymnastics.com>
- Fédération Internationale de Gymnastique. (2005). *2005 FIG judges' rules*. Retrieved June 8, 2006, from <http://www.fig-gymnastics.com/index2.jsp?menu=RULES>
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117–140.
- Geiselman, R. E., Haight, N. A., & Kimata, L. G. (1984). Context effects on the perceived physical attractiveness of faces. *Journal of Experimental Social Psychology*, 20, 409–424.
- Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science*, 5, 152–158.
- Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, 52, 45–56.
- Gilbert, D. T., Giesler, R. B., & Morris, K. A. (1995). When comparisons arise. *Journal of Personality and Social Psychology*, 69, 227–236.
- Gilovich, T. (1984). Judgmental biases in the world of sports. In W. F. Straub & J. M. Williams (Eds.), *Cognitive sport psychology* (pp. 31–41). New York: Sport Science Associates.
- Helson, H., & Kozaki, A. (1968). Anchor effects using numerical estimates of simple dot patterns. *Perception and Psychophysics*, 4, 163–164.
- Herr, P. M. (1986). Consequences of priming: Judgment and behavior. *Journal of Personality and Social Psychology*, 51, 1106–1115.
- Herr, P. M., Sherman, S. J., & Fazio, R. H. (1983). On the consequences of priming: Assimilation and contrast effects. *Journal of Experimental Social Psychology*, 19, 323–340.
- Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York: Guilford Press.
- Higgins, E. T., & Lurie, L. (1983). Context, categorization, and recall: The “change-of- standard” effect. *Cognitive Psychology*, 15, 525–547.
- Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, 13, 141–154.
- Hovland, C., Harvey, O., & Sherif, M. (1957). Assimilation and contrast effects in reactions to communication and attitude change. *Journal of Abnormal and Social Psychology*, 55, 244–252.
- Jarvis, B. (2002). MediaLab, Direct RT [Computer software]. New York: Empirisoft Research Software.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Kenrick, D. T., & Gutierrez, S. E. (1980). Contrast effects and judgments of physical attractiveness: When beauty becomes a social problem. *Journal of Personality and Social Psychology*, 38, 131–140.



- Kernis, M. H., & Wheeler, L. (1981). Beautiful friends and ugly strangers: Radiation and contrast effects in perception of same-sex pairs. *Personality and Social Psychology Bulletin*, 7, 617–620.
- King, D. L. (1988). Assimilation is due to one perceived whole and contrast is due to two perceived wholes. *New Ideas in Psychology*, 6, 277–288.
- King, M. C., & Lockhead, G. R. (1981). Response scales and sequential effects in judgment. *Perception and Psychophysics*, 30, 599–603.
- Krantz, D. L., & Campbell, D. T. (1961). Separating perceptual and linguistic effects of context effects upon absolute judgment. *Journal of Experimental Psychology*, 62, 35–42.
- Landers, D. M. (1970). A review of research on gymnastic judging. *Journal of Health Physical Education and Recreation*, 41, 85–88.
- Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology*, 73, 91–103.
- Manis, M. (1971). Context effects in communication: Determinants of verbal output and referential decoding. In M. H. Appley (Ed.), *Adaptation-level theory: A symposium* (pp. 237–255). New York: Academic Press.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100, 254–278.
- Mellers, B. A. (1986). “Fair” allocations of salaries and taxes. *Journal of Experimental Psychology: Human Perception and Performance*, 12, 80–91.
- Morse, S., & Gergen, K. J. (1970). Social comparison, self-consistency, and the concept of self. *Journal of Personality and Social Psychology*, 16, 148–156.
- Mussweiler, T. (2001a). Focus of comparison as a determinant of assimilation versus contrast in social comparison. *Personality and Social Psychology Bulletin*, 27, 38–47.
- Mussweiler, T. (2001b). “Seek and ye shall find”: Antecedents of assimilation and contrast in social comparison. *European Journal of Social Psychology*, 31, 499–509.
- Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review*, 110, 472–489.
- Mussweiler, T. (in press). Assimilation and contrast as comparison effects: A selective accessibility model. In D. A. Stapel & J. Suls (Eds.), *Assimilation and contrast in social psychology*. New York: Psychology Press.
- Mussweiler, T., & Bodenhausen, G. (2002). I know you are but what am I? Self-evaluative consequences of judging ingroup and outgroup members. *Journal of Personality and Social Psychology*, 82, 19–32.
- Mussweiler, T., & Englich, B. (2005). Subliminal anchoring: Judgmental consequences and underlying mechanisms. *Organizational Behavior and Human Decision Processes*, 98, 133–143.
- Mussweiler, T., Rüter, K., & Epstude, K. (2004a). The man who wasn’t there: Subliminal social comparison standards influence self-evaluation. *Journal of Experimental Social Psychology*, 40, 689–696.
- Mussweiler, T., Rüter, K., & Epstude, K. (2004b). The ups and downs of social comparison: Mechanisms of assimilation and contrast. *Journal of Personality and Social Psychology*, 87, 832–844.
- Mussweiler, T., & Strack, F. (2000). The “relative self”: Informational and judgmental consequences of comparative self-evaluation. *Journal of Personality and Social Psychology*, 79, 23–38.
- Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and short-comings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.
- Nosanchuk, T. A., & Erickson, B. H. (1985). How high is up? Calibrating social comparison in the real world. *Journal of Personality and Social Psychology*, 48, 624–634.
- O’Brien, K. (1991). Bias in the judging of international elite gymnasts. In J. Standeven, K. Hardman, & D. Fisher (Eds.), *Sport for all: Into the 90s* (pp. 148–153). Aachen, Germany: Meyer & Meyer.
- Ostrom, T. M., & Upshaw, H. S. (1968). Psychological perspectives and attitude change. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes* (pp. 217–242). New York: Academic Press.
- Parducci, A. (1963). Range–frequency compromise in judgment. *Psychological Monographs*, 77(2, Whole No. 565).
- Parducci, A. (1965). Category judgment: A range–frequency model. *Psychological Review*, 72, 407–418.
- Parducci, A. (1968). The relativism of absolute judgment. *Scientific American*, 219, 84–90.
- Parducci, A., & Marshall, L. (1962). Assimilation versus contrast effects of anchoring stimuli on judgments. *Journal of Experimental Psychology*, 63, 426–437.
- Parducci, A., Perrett, D. S., & Marsh, H. W. (1969). Assimilation and contrast as range–frequency effects of anchors. *Journal of Experimental Psychology*, 81, 281–288.
- Parducci, A., & Wedell, D. H. (1986). The category effect with rating scales: Number of categories, number of stimuli, and method of presentation. *Journal of Experimental Psychology: Human Perception and Performance*, 12, 496–516.
- Pelham, B. W., & Wachsmuth, J. O. (1995). The waxing and waning of the social self: Assimilation and contrast in social comparison. *Journal of Personality and Social Psychology*, 69, 825–838.
- Pepitone, A., & DiNubile, M. (1976). Contrast effects in judgments of crime severity and the punishment of criminal violators. *Journal of Personality and Social Psychology*, 33, 448–459.
- Plessner, H. (1997). *Urteilsverzerrungen bei Kampfrichtern im Kunstturnen—Der Einfluß von Erwartungen* [Judgment biases of gymnastic judges—The influence of expectancies]. Aachen, Germany: Shaker.
- Plessner, H. (1999). Expectation biases in gymnastics judging. *Journal of Sport and Exercise Psychology*, 21, 131–144.
- Plessner, H. (2005). Positive and negative effects of prior knowledge on referee decisions in sports. In T. Betsch, & S. Haberstroh (Eds.), *The routines of decision making* (pp. 311–324). Hillsdale, NJ: Erlbaum.
- Plessner, H., & Haar, T. (in press). Judging sport performance from a social cognition perspective. *Psychology of Sport and Exercise*.
- Plessner, H., & Schallies, E. (2005). Judging the cross on rings: A matter of achieving shape constancy. *Applied Cognitive Psychology*, 19, 1145–1156.
- Salmela, J. H. (1978a). Gymnastics judging: A complex information processing task, or (who’s putting one over on who?) Part 1. *International Gymnast*, 20, 54–56.
- Salmela, J. H. (1978b). Gymnastics judging: A complex information processing task, or (who’s putting one over on who?) Part 2. *International Gymnast*, 20, 62–63.
- Scheer, J. K. (1973). Effect of placement in the order of competition on scores of Nebraska high school students. *Research Quarterly*, 44, 79–85.
- Scheer, J. K., & Ansorge, C. J. (1975). Effects of naturally induced judges’ expectations on the ratings of physical performances. *Research Quarterly*, 46, 463–470.
- Scheer, J. K., Ansorge, C. J., & Howard, J. (1983). Judging bias induced by viewing contrived videotapes: A function of selected psychological variables. *Journal of Sport Psychology*, 5, 427–437.
- Sherif, M., Taub, D., & Hovland, C. I. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of Experimental Psychology*, 55, 150–155.
- Sigall, H., & Landy, D. (1973). Radiating beauty: Effects of having a physically attractive partner on person perception. *Journal of Personality and Social Psychology*, 28, 218–224.
- Smith, E. R., & Zárate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review*, 99, 3–21.
- Sport 1. *The development of the world record in ski jumping*. (2006). Retrieved June 10, 2006, from <http://sport1.at/coremedia/generator/id=1965212.html>

- Stapel, D. A., & Suls, J. (2004). Method matters: Effects of explicit versus implicit social comparisons on activation, behavior, and self-views. *Journal of Personality and Social Psychology*, 87, 860–875.
- Stefani, R. (1998). Predicting outcomes. In J. Bennett (Ed.), *Statistics in sport* (pp. 249–275). London: Arnold.
- Ste-Marie, D. M., & Lee, T. D. (1991). Prior processing effects on gymnastic judging. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 126–136.
- Ste-Marie, D. M., Valiquette, S. M., & Taylor, G. (2001). Memory-influenced biases in gymnastic judging occur across different prior processing conditions. *Research Quarterly for Exercise and Sport*, 72, 420–426.
- Strack, F. (1992). The different routes to social judgments: Experiential versus informational strategies. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgment* (pp. 249–275). Hillsdale, NJ: Erlbaum.
- Suls, J., Gastorf, J. W., & Lawhon, J. (1978). Social comparison choices for evaluating a sex- and age-related ability. *Personality and Social Psychology Bulletin*, 4, 102–105.
- Upshaw, H. S. (1969). The personal reference scale: An approach to social judgment. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 4, pp. 316–371). New York: Academic Press.
- Ward, L. M., & Lockhead, G. R. (1971). Response system processes in absolute judgment. *Perception and Psychophysics*, 9, 73–78.
- Wedell, D. H., & Parducci, A. (1988). The category effect in social judgment: Experimental ratings of happiness. *Journal of Personality and Social Psychology*, 55, 341–356.
- Wedell, D. H., & Parducci, A. (2000). Social comparison: Lessons from basic research on judgment. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 253–270). New York: Plenum.
- Wedell, D. H., Parducci, A., & Geiselman, R. E. (1987). A formal analysis of ratings of physical attractiveness: Successive contrast and simultaneous assimilation. *Journal of Experimental Social Psychology*, 23, 230–249.
- Wheeler, L. (1966). Motivation as a determinant of upward comparison. *Journal of Experimental Social Psychology*, 2, 27–31.
- Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin*, 90, 245–271.
- Wilson, V. E. (1976). Objectivity, validity, and reliability of gymnastic judging. *Research Quarterly*, 47, 169–174.
- Wyer, R. S., & Srull, T. K. (1989). *Memory and cognition in its social context*. Hillsdale, NJ: Erlbaum.

Received February 9, 2006

Revision received June 27, 2006

Accepted June 27, 2006 ■

## Low Publication Prices for APA Members and Affiliates

**Keeping you up-to-date.** All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*. High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

**Essential resources.** APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association*, and on dozens of new topical books each year.

**Other benefits of membership.** Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

**More information.** Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.