From role conflict to evaluation discordance:
How do conflicting performance evaluations affect risk taking in multiple audience contexts?

(Chapter for inclusion in Svyantek and Mahoney, Eds., "Received Wisdom, Kernels of Truth, and Boundary Conditions in Organizational Studies")

Edward Bishop Smith
University of Michigan

July 2012
Title: From role conflict to evaluation discordance: How do conflicting performance evaluations affect risk taking in multiple audience contexts?

Abstract: A rich but checkered literature in sociology presumes that maintaining multiple and conflicting social roles—for instance, having a high level of education and low level of occupational prestige—produces anxiety. Due to many theoretical and methodological difficulties involved with identifying the source and consequences of these "role conflicts," however, such theories have been largely neglected in contemporary discourse. In this chapter I revisit this influential but discredited line of research in order to combine its most basic intuition—social life involves living in multiple contexts simultaneously—with several insights gleaned from more enduring research on the sociology and social psychology of identity. Ultimately, I propose an alternative to earlier role conflict theories: anxiety, and certain behavioral responses intended to alleviate that anxiety, are not the product of conflicting roles, per se, but of conflicting evaluations by two or more discrete and relevant audiences of a single role performance. I evaluate this hypothesis using panel data from the Professional Golf Association (PGA). Results show that golfers are more likely to adopt risky playing strategies when a discontinuity exists between their internally- (players) and externally- (media) derived evaluations. Implications for social comparison and performance feedback theories under conditions of multiple audiences are highlighted.
**Introduction**

Individuals and organizations alike often play multiple social roles. An adult male, for instance, may play employee, father, and husband (and possible several other things) within the span of a single day. A firm can play market leader in an industrial context and concerned community member in a social one. Although some roles may never come in contact with one another, others can carry competing expectations and present conflicts for those occupying them. Indeed, it was precisely the observation that social actors are obliged to simultaneously play multiple, possibly conflicting roles that motivated a significant amount of research in sociology that investigated the consequences of "role conflicts" and "status inconsistencies" (Benoit-Smullyan 1944; Hughes 1944; Lenski 1954, 1956; Marks 1977). According to these theories, people whose status is higher in one role than it is in another will experience frustration and anxiety, and that this emotional response will subsequently affect both preferences and behaviors. Despite the initial appeal of such theories, however, they are now largely overlooked due to an array of theoretical and methodological problems that came to be associated with the early work. In this chapter I revisit role conflict theory in order to integrate one of its core assumptions—i.e., social life involves residing in multiple contexts simultaneously—with more enduring theories about social comparison and performance feedback (e.g., Festinger, 1954; Cyert and March, 1963; Greve, 2003). In their integration, a new variant of an old question emerges: What is the effect of receiving contrasting evaluative feedback from two or more relevant audiences? Furthermore, when multiple evaluations of the same role-performance differ, how is the actor being evaluated likely to respond to the conflicting evaluations?

There are at least three compelling reasons for summarizing and integrating theories about role conflict with research on social comparison and performance feedback. First, like each of the latter theories, role conflict theories were originally developed to explain psychological anxiety and account for social behaviors meant to alleviate that anxiety. Second, what each theory lacks the others offer. To role conflict theories, social comparison and performance feedback highlight the
importance of external reference groups and evaluations. To social comparison and performance feedback, role conflict theories, with their emphasis on the multiplexity of social life, can motivate an expanded search for relevant sources of comparison and feedback. Finally, as role conflict theories were partially undone by their strong dependence on context—even if an African American doctor felt immense anxiety in the 1950s and 1960s (e.g., Lenski 1954), is that same doctor as likely to feel that anxiety now, or in a more progressive community?—integration with some of the more generalizable tenets of social comparison theories may breathe new life (i.e., generalizeability) into a largely overlooked field of research.

At the heart of early role conflict theories is an intuition that is as powerful as it is simple. Regular, automatic behaviors may very well conform to learned scripts (Mische and White 1998) and routines (DiMaggio and Powell 1983), but purposeful action occurs at the intersections of social life. Role conflict theories are thus uniquely positioned to highlight the sources of purposeful actions. Meaningful, conscious action materializes at the confluence of conflicting social roles. Indeed, the existence and implications of role conflicts can be found far beyond social science texts. The author Charlotte Bronte’s character, Jane Eyre, offers an example. Eyre is the embodiment of the conflicting roles that characterized Victorian governesses. As both relevant and sophisticated members of the upper echelon but nonetheless paid employees of it, the antagonistic nature of governesses' conflicting roles are pointed to by historians and literary critics alike as the cause of severe mental anguish among this class of women (Peterson 1970).

Coupling the intuitive appeal of role conflict theories—which I explore in more detail in the next section of text—with what I will highlight to be a weak empirical and methodological history, I suggest in this chapter a shift in attention from multiple roles to multiple perceptions of a single role. To motivate this shift I turn again to the humanities, though this time not literature, but art: critics speculate that the faces in expressionist artist Edvard Munch’s, “The Anxiety,” are meant to capture the melancholy and stress of individuals who are viewed in one way by the world in which they live
(the painting) and another way by the world outside (the very world peering in at them). Anxiety here is not the product of maintaining multiple roles, but rather being evaluated differently by two audiences for the same role performance. My argument is this; whereas discrete, even conflicting roles may be meaningfully decoupled (White, 2008; Bothner, Smith, & White, 2011) and prioritized (e.g., Coser 1974), multiple and contradictory evaluations of a single role performance by two or more relevant audiences may not. In much the same way that people respond positively to sequences of related words (and negatively to sequences of unrelated words) (Kahneman 2011), I argue that people are systematically prone to feel some measure of anxiety when two or more evaluations of a single role performance do not align. Being praised by colleagues but overlooked by upper management, for instance, means having to process two discrete sources of feedback that are intended to describe the same role performance. Anxiety occurs at the confluence of misaligned perceptions when those perceptions are of an actor’s performance in a single role.1

In what follows I situate the proposed theoretical shift—from role conflict to evaluation discordance—more intimately into the sociological and social psychological literatures. I suggest that for any given role performance, more than one evaluation may materialize to describe that performance. In other words, social actors are subject to multiple streams of performance feedback (cf., Cyert and March 1963; Greve 2003). For the sake of my empirical analysis, I focus on two different types of evaluations: one emanating from others who are vying for identical resources—i.e.,

1 Determining what constitutes a single role is not a straightforward task. All social roles, it seems, are regresable to the point of the solitary ego (and perhaps further if one considers cognitive divisions). When we consider “role” from the perspective of the observer, the “employee” mentioned in the text plays two roles, “colleague” amongst peer employees and “subordinate” to upper management. This possibility is seemingly at odds with my discussion of an individual being perceived differently in a single role. Two things help to mitigate the apparent contradiction. First, whereas colleague and subordinate might be decoupled as discrete roles, it is more difficult to decouple an individual’s performance under the broader role-heading, employee. Second, while I do not do it in this paper, one may opt to relax the single role assumption and consider the relationship between role distance (colleague and subordinate are closer than colleague and congregation member, for instance, because they fall under the same umbrella role, employee) and the likelihood of evaluation decoupling. As the distance between two roles increases, they should be more easily decoupled from one another and evaluation discordances should not cause strain. When the distance is more minute, however, misalignments in audience perceptions increasingly weigh on the individual, producing strain and resulting in behavioral changes meant to alleviate that strain.
competitors—and a second from more distant, but nevertheless relevant, onlookers—i.e., analysts, commentators, audiences, etc. The former constitutes an "insiders" or proximate source of feedback. The latter, by comparison, is an "outsiders" or distant source of feedback. Receiving positive proximate feedback describes the highly respected colleague or dominant competitor. Positive distant feedback, by contrast, might best characterize a headlining CEO, publicly praised academic, or idolized champion. The correlation between proximate and distant feedback should be significant in many cases, but need not be in all.

Interlaced in the more detailed theoretical discussion to follow, I also explore the principal methodological implications of the shift from role conflict to evaluation discordance, of which there are three. First, how should we identify multiple and discrete audiences and the feedback they generate? Second, how should we specify a discordance between that feedback? And last, how might institutional context determine the effect of a discordance?

Theoretical Antecedents

An appreciation of people's capacity to play multiple social roles has long been a part of sociological discourse from Weber (1978) to Parsons (1953) on through to more modern discussions of role and identity (Easterlin 1980; Kogut and Zander 1996; Uhlenberg and Riley 1996). Most discussions of role multiplexity, whether directly or indirectly, draw on Simmel’s (1955) description of social structure as a complex of crosscutting social circles and memberships. Indeed, Simmel's imagery of intersecting affiliations—expanded upon by Blau and Schwartz (1984)—has influenced a wide range of social theorists. Coser (1974), for one, argued that “people occupy not just one status, but a complex of distinct status positions. For example, they are at the same time, placed in the stratification system, the marital order, the division of labor, the religious order.” Much like role conflict theories of the time, Coser considered the potential for dysfunction via incompatible roles: when car salesmen sell to a friend, for instance, do they play the role of friend by conceding to bargain pricing or play salesman and attempt to maximize profit on behalf of themselves and the
dealership? While such conflicts constitute the origin of tension and anxiety in many role conflict hypotheses (and demand subsequent behavior to relieve such tension), for Coser specifically, tension was avoided according to his assumption that people have an ability to prioritize their goals, and therefore their roles, at any given point in time.

The conflict Coser resolved through goal prioritization are left unresolved by a second, related account of role conflict. In fact, the tension between incompatible roles is the key point of empirical significance in status inconsistency theory. Conceptually introduced by Benoit-Smullyan (1944) and Hughes (1944) and articulated and coined by Lenski (1954, 1956), status inconsistency theory maintains that a significant amount of information is embedded in the (in)consistency among an individual’s positioning across various status hierarchies, namely income, occupation, education, and ethnicity. According to Lenski a person positioned at comparable levels across each of the vertical dimensions is status consistent, or "status crystallized." Status inconsistencies matter, according to Lenski, not simply because they generate anxiety in the person who is status inconsistent, but also because inconsistencies are important predictors of people's preferences and behaviors. In a study of Detroit residents, Lenski argued, “individuals characterized by a low degree of status crystallization differ significantly in their political attitudes and behavior from individuals characterized by a high degree of status crystallization, when status differences in the vertical dimensions are controlled” (1954, p. 405). Specifically, Lenski showed that people marked by low crystallization maintained liberal stands on controversial political issues.

Following Lenski, a number of studies emerged citing status inconsistency as an important cause of health outcomes (Jackson 1962), political behavior (Goffman 1957; Coleman 1957, Bell 1964), and feelings of injustice (Bettelheim and Janowitz 1964). As the amount of research in support of status inconsistency and role conflict theories grew, however, a second stream of research emerged in the 1960s and 1970s questioning both the theoretical and methodological merit of the approach (e.g., Blocker and Riedesel 1978; Crosbie 1979; Whitt 1983). Among these, Treimen
(1966) argued that the resulting interaction or inconsistency term has little to no explanatory power after baseline statuses are properly controlled. Bauman (1968) focused on the high level of context dependence inherent to role conflict and status inconsistency theories by illustrating that the negative effects of inconsistency in several of Lenski’s earlier models should be positive in alternative empirical settings. Laumann and Segal (1971) offered an even more direct challenge of Lenski’s thesis, suggesting instead that the persistence of traits that are characteristic of certain subcultures offer a more parsimonious explanation of Lenski’s primary dependent variables. In other words, like Bauman, Laumann and Segal found that more properly specified main effects trumped the proposed effects of the inconsistency interaction. Finally, Meyer and Hammon (1971) challenged the framework on more theoretical grounds by posing three important questions. First, what combination of attributes or statuses should be taken into account to determine an inconsistency or crystallization? Second, why should inconsistencies result in anxiety, specifically? And third, why should inconsistencies matter at all?

Towards Evaluation Discordance

Meyer and Hammon’s (1971) first critique of status inconsistency asks what combination of attributes or statuses should be accounted for to determine a crystallization or inconsistency. Determining what roles to include and what roles to drop from an analysis is seldom straightforward. Furthermore, it is unclear when an effect can be interpreted as representative of a generalizable, behavioral mechanism and when it is primarily the result of context. To revisit an example from above, although there is nothing innately stressful about being a physician and an African American, in some contexts the combination might elicit anxiety. It is not obvious that an individual should worry about a portfolio of roles, at least until such a time when two or more of those roles are proven to be both relevant and contrasting.

In comparison to role conflicts, I argue that being perceived or evaluated differently by two or more relevant audiences for the same role performance is likely to be discomforting independent
of context. In other words, my hypothesis is about multiple and contemporaneous evaluations of the same person playing the same role. Although several related theses can be found in prior research on social comparison, none seem to fit this proposition explicitly. For instance, Davies (1969) and Gurr (1970) found violence and social strife to be the result of conflicting outcomes experienced by the same person but at different points in time—usually comparing the present to the past (for a review of this and related research, see Levine and Moreland, 1987). Folger and colleagues (1983) extended the notion of within-person social comparison to include conflicting and contemporaneous outcomes (like I am proposing here), but focused their attention on comparing real-world outcomes or evaluations to hypothetical ones that people imagine according to counter-factual logic. My focus is on intrapersonal, conflicting, and contemporaneous outcomes—feedback or evaluations, specifically—that originate from multiple audiences but are intended to describe the same thing—an actor's performance in a single role.

The shift in reasoning from conflicting roles to conflicting evaluations mitigates to some extent Meyer and Hannom’s second critique of role conflict and status inconsistency theories—why should inconsistencies result in anxiety in the first place? While there is indeed no systematic evidence to suggest that maintaining conflicting roles will reliably produce anxiety (Bauman, 1968), there is a wide range of empirical evidence supporting the notion that people prefer consistent views and evaluations of themselves—for instance, preferring people who have views of them that match their own views (Swann, 1983). People experience dissonance or discomfort when faced with two inconsistent cognitions (Ibarra & Barbulescu, 2010; Ibarra 1999; Ebaugh, 1988; Festinger, 1954), as would be generated by differing feedback or evaluations from two or more relevant audiences. Consistent feedback—whether consistently negative or consistently positive (e.g., Swann, Wenzlaff, & Tafarodi, 1992)—leads people to feel more comfortable.

To illustrate further the importance and implications of consistent or inconsistent evaluations,
consider the following two anecdotes. Anecdote one: the high rate of suicide among dental professionals has been attributed to high levels of stress and anxiety that result from the profession having positive in-group perception but low public perception relative to other medical professionals (Stack 1996). Anecdote two comes from an organizational context, but speaks to an equivalent desire for consistent feedback: Proctor & Gamble's mission statement reads; “Be, and be recognized as, the best consumer products services company in the world” (emphasis added). An internal, or proximate assessment of "being the best" should come coupled with a similar assessment from a less proximate audience (such as customers or industry analysts). Being positively evaluated by a set of insiders—whether confederate dentists or managerial fiat—is not enough. As Goffman highlighted, tension is the result of a sensed discrepancy “between the world that spontaneously becomes real to the individual and the one in which he is obliged to dwell” (1961, p. 43). It may not be a second role after all, but precisely a second audience that sets the stage for conflict.

What kind of behavior should we expect to be associated with conflicting performance evaluations? To answer this question I draw on the behavioral theory of the firm (March and Simon, 1958; Cyert and March, 1963; see also Greve, 2003). The link is simple: according to the behavioral theory of the firm, managers use social comparison to evaluate their own performance against that of some reference group. When performance falls short of the reference group, managers commence "problemistic search" designed not only to identify the cause of the performance deficit, but also to make changes (oftentimes risky ones) in anticipation of eliminating it. I expect that conflicting performance evaluations work in much the same way. When people receive contrasting feedback from two or more relevant and discrete audiences for the same role performance, they are likely to engage in problemistic search as a way to account for the discrepancy and possibly bring the two evaluations into alignment.

To recap, the array of criticisms of role conflict theories may come down to one thing: perhaps Coser was right about the decoupling capacity of the human psyche. By prioritizing one’s
goals at any one point in time, a person may subvert the rise of anxiety that would otherwise be associated with maintaining conflicting roles. Lisa Leslie, an American women's basketball player, spoke to this effect when recently asked how she handles being a mother and an Olympic athlete, Leslie answered by saying that she is never more than one at a time; “When I’m with [my daughter], I am 100% mother. When I’m on the court, I’m 100% basketball player” (Leslie, 2008).

Accordingly, if Leslie were criticized for her playing, it should have little effect on her self-perception as a mother, as vice versa. I argue that such decoupling is more difficult, however, when two or more evaluations of an actor are meant to describe the same role. An academic whose work is received well by one audience and poorly by another is apt to be affected by the discrepancy, so long as she considers both audiences to be equally important. A politician deemed visionary by Washington pundits and washed-up by a home constituency may become anxious as a result of the misalignment—perhaps fearing that the positive evaluation of one audience will be affected by the negative evaluation of the other—and engage in search designed to eliminate the inconsistency.

**Conceptualizing Multiple Audiences and Observing the Effects of Evaluation Discordance**

An important piece of this chapter's core proposition that remains to be explained is how inconsistent evaluations might arise in the first place. Earlier role conflict theories were mostly exempt from addressing this sort of question. Given some existing and shared understanding of a status ordering—such as Duncan’s (1984) scale of occupation prestige—demonstrating how status inconsistencies emerged constituted little more than a search for discrepant data points. The complexity of social life in an environment where people are (mostly) free to make their own decisions ensures that at least some people will exist off the diagonal. For me, the question is more complex. Why should two audiences perceive an actor or an actor’s performance in a given role in starkly different ways? Surely misalignments may happen by chance. A more systematic answer to this question should depend, however, on two things. The first is ambiguity in evaluative criteria. When a role performance is judged using objective metrics, there should be less room for two
audiences to make evaluations that are misaligned, compared to when subjective metrics are the norm for evaluation. The second factor has to do with the composition of the audience groups themselves. Other things held constant, when two audiences are composed of individuals who are equally knowledgeable about the thing they are evaluating, there should be a high level of evaluative agreement between the groups and little opportunity for discordance. If one of the audiences is composed of people who vary in their knowledge, however, that audience may be forced to employ less nuanced metrics—a kind of "ratcheting down"—by which to discuss and evaluate a given actor's role performance. In this case, because the two audiences use different criteria to converse and make evaluations, discrepancies between them are more probable.

The preceding discussion brings to light one additional question: what constitutes discrete audiences? In addressing this question, I advocate starting with a single distinction; "us" and "them," or more generally, "inside" and "outside." While each of these distinctions may be further parsed, for my purposes here the simple dichotomy proves sufficient. Moreover, there is considerable precedent for exploiting a basic insider-outsider distinction, as it constitutes an important theoretical construct in a number of empirical and ethnographic accounts of social change, coordinated action, and individual perception and behavior. The dichotomy describes Merton’s (1957) local and cosmopolitan (see also Gouldner 1957, 1958), Gouldner’s (1954) indulgent and productive, and Abbott’s (1981) intraprofessional and public. Consequently, the audience identification strategy that I propose separates feedback into two types, proximate and distant, by using a focal actor’s orientation to resources to determine what kinds of feedback should be deemed as one kind versus another. Evaluations originating from actors who are vying for the same resources as the focal actor constitute proximate feedback. Evaluations originating from actors who rely on a different resource pool than the focal actor form distant feedback. In the next section I describe how I apply this distinction to identify two discrete and relevant "audiences" in the context of professional golf. Once audiences are determined and feedback is measured, I use the context to test the evaluation discordance hypothesis.
Methods

Empirical Setting: The Professional Golf Association (PGA)

Several considerations drew me to professional golf as an appropriate study context. I will highlight two here. First, there are (at least) two identifiable and discrete sources of performance feedback in the world of professional athletics: one is simply the group of other competitors or peers, another is media commentators and the general public. As I will describe in more detail below, the evaluations associated with both of these groups are observable and, more importantly, measureable. Second, there is a plethora of data collected and maintained by the PGA that can be used to construct a variety of important predictor and outcome variables for identification and analysis of evaluation discordance.

The PGA itself is composed of three separate tours—the PGA Tour (the primary professional golf tour), the Champions Tour (a tour open to golfers 50 years of age and older), and the Nationwide Tour (a “minor league” tour commonly regarded as a feeder for the PGA)—and operates more than 100 yearly tournaments. Of these 100, roughly half are classified as PGA Tour events and combine to make up a total purse of over $250,000,000 in prize winnings. I collected data for the PGA Tour only and confine my analyses to events in which four rounds of play were completed. Furthermore, due to a major shift in the competitive format of the PGA beginning with the 2007 season, I use only data up to and including the 2005 season. According to these considerations and restrictions, the final dataset consists of results from 223 total events, or approximately 60,000 player-round observations, spanning the 2001 through 2005 seasons.

---

2 There is one 5-round tournament on the PGA Tour that was dropped from the analysis. In addition, those events that were shortened due to weather conditions are not included in the analysis.
3 In 2007 the PGA instituted the PGA FedEx Cup, a year-long points-based competition to award the season’s best golfer. In addition to changing the order of several PGA events, under the new format the season ends with a playoff-like schedule for which players must qualify. It is reasonable to think that this shift impacted individuals’ playing strategies in 2007 – playing conservatively in order to qualify for the playoff as opposed to playing only to win, for instance – and that players may have used the 2006 season as a test season for strategic changes.
The PGA Tour season extends from January to November. With a few notable exceptions, each tournament is made up of four rounds and is played over a four-day period, typically extending from an opening round on Thursday to a closing round on Sunday. Player participation hinges on a series of exemptions and qualifications and is ultimately determined by the player. In an average year, the average PGA Tour player is likely to play in approximately 25 events, with some players opting to play more and some choosing or being able to play less. A field of 150 golfers enters most PGA Tour events and completes the first two rounds of play over a two-day period. Following the close of the second round of play, the 70 players with the lowest (best) scores, including ties, advance to the final two rounds of play and are guaranteed a percentage share in the total prize money. While the total amount of prize money awarded varies by tournament (largely as a function of the prestige of the tournament), allocation of the prize is determined by a standard PGA Tour formula and is characterized by a convex payout distribution.

**Dependent Variable: Risk**

I hypothesized that receiving misaligned or discordant evaluations by two or more relevant audiences should increase player's level of anxiety and trigger problemistic search (Cyert and March 1963; Greve 2003). Like most studies that rely on archival data there is naturally no way given the available data to observe anxiety. Instead, I rely on an observable, behavioral outcome that past research has linked explicitly to problemistic search: risk taking (Greve, 2003; Baum and Dahlin, 2007; Kacperczyk, Beckman, & Moliterno, 2011). I measure risk taking behavior in golf as the number of times in a round a player "goes for the green," or makes an attempt to hit the ball onto the green on a player’s tee (first) shot for par 4 holes or second shot for par 5 holes. Due to the length and difficulty of this kind of shot, "going for the green" is commonly regarded as an indicator of risky play (see Pope and Schweitzer 2011 for a related, putting-specific measure). Risk enters the predictive models described below as a yearly average, ranging from a minimum of 0.26 to a maximum of 2.53 and having a mean of 1.32.
Independent Variables: Performance Feedback, Proximate and Distant

The first feedback measure, a golfer’s proximate feedback, is constructed using data on golfers’ finishing ranks in previous tournaments. I define a player with high proximate feedback as one who dominates other players who in turn dominate others (Ridgeway and Walker 1995; Bothner, Kim, and Smith 2012). In other words, I model proximate feedback as a function of being favorably situated in a pecking order (Chase 1980). My measure of proximate feedback extends from prior sociological research on competition and localized status. For instance, Park (1952: 574) argued that, “the status of the individual…is determined by rivalry.” Likewise, Blau (1964: 127) theorized, the “differentiation of social status…emerges in the course of competition [as] each group member competes with all the other members for [their] respect.” In more contemporary literature, Podolny (1993; 2005) measured status among investment banks according to banks’ dominant positions over other banks in investment banking deals. Related examples include the athlete who outperforms other top athletes, the juvenile who intimidates other prominent juveniles (e.g., Whyte 1943), and the scientist who is recognized by other elite scientists.

When computing proximate feedback, I start by assembling for each tournament an asymmetric matrix $D_t$. Within this matrix, cell $d_{ij}$ tallies, over a one-year moving window, the number of ranks by which $i$ outperformed $j$ during those events in which $i$ came in ahead of $j$. Consequently, if during the one-year moving window, there were three tournaments in which $i$ placed ahead of $j$, and in those three events, $i$ did so by a total of ten ranks (perhaps by five ranks in the first, four ranks the second, and one rank in the third), then the cell marked by row $i$ and column $j$ equals ten. Similarly, $d_{ji}$ reports, again over a one-year moving window, the number of ranks by which $j$ outstripped $i$, thus introducing the asymmetry.
Using Bonacich’s (1987) measure of network power, I then compute proximate feedback scores as: \( P_{it}(\alpha, \beta) = \sum_j (\alpha + \beta P_{jt}) d_{ij} \), where \( P_{it} \) is an element of vector \( P_t \) denoting the proximate feedback of actor \( i \) at time \( t \). I select the scaling parameter \( \alpha \) so that, in each period, regardless of the size of the population, the actor whose entry equals 1 in \( P_t \) does not possess a disproportionately high or low level of feedback (Bonacich 1987, p. 1173). I set the parameter \( \beta \) equal to \( \frac{3}{4} \) of the reciprocal of the norm of the maximum eigenvalue of \( D \), following prior studies (Bothner, Kim, and Smith 2012; Bothner, Smith, and White, 2011). To the extent that golfers enjoy high values in \( P_t \), they have recently outranked the highest-ranking athletes in the sport.

The second evaluation, a player’s distant feedback, is simpler in construction. For each week in the panel, I tally the number of times a player is mentioned in a sporting print media outlet. There are three primary reasons why a sports media-based measure appropriately captures distant feedback, or that component of a player’s aggregate feedback which is derived from a relevant, but non-competitor audience. First, the measure has built into it a control for the number of rounds played by each golfer. Some media mentions, indeed the majority of mentions for some less prestigious players, are nothing more than a box-score, "player did compete" line. Second, while I do not suppose that every golfer is acutely aware of how many times he is mentioned by sports reporters, the correlation between this count measure and a player’s prominence among a general golfing audience is likely to be high. Third, though some sporting media outlets do occasionally choose to highlight an athlete for reasons other than his athletic performance, I expect these occasions to be less frequent than if non-sporting media were included. This is important as the proposed hypothesis assumes the generation of anxiety to occur when an actor receives differing feedback for the same role performance. Tour player Bubba Watson’s above average play should be no cause for panic, for instance, when contrasted with media jests about his now infamous lack of fashion sense. When the sporting media
overlooks him despite his strong playing record, however, this, I contend, constitutes a source of insecurity.

As I discussed above, one of the primary problems that came to be associated with role conflict theories and status inconsistency, specifically, was that there was no accepted method by which to measure the effect of status discordances. The earliest approaches were sharply criticized for failing to include main effects in their models. As subsequent approaches addressed these criticisms by including main effects, there continued to be no agreement on how an inconsistency should be measured. Interactions failed on account of interpretability. Dummy indicators of inconsistency proved simpler from an interpretive standpoint but suffered from a lack of clarity. Like the interaction-based approach, determining what did and did not constitute an inconsistency was left purely to the discretion of the analyst. More systematic attempts redeployed one of the earliest approaches, standardizing variables to generate equivalent distributions and then calculating inconsistencies as differences between one's place in each of the new distributions (Hartman 1974). Doing so, of course, returned things right back to the problem of adequately specifying main effects—two main effects and a difference between them amounts to an over-specified model.

Fortunately, the structure of my data allows me to employ a more straightforward method in this chapter. By exploiting the longitudinal nature of the PGA panel, I am able to use standard OLS estimation to capture the non-additive evaluation discordance effect over time. Specifically, I parse the distribution of distant feedback scores into three categories: low, high, and very high. I opt for the term "very high," as opposed to using the more conventional low, medium, and high, for two reasons. The first is purely technical and captures the right-skewness of the underlying distribution of media mentions—median = 36, mean = 79.5, max = 1000. The second is more conceptual: Goffman offers reason to believe that a relatively small number of people may be relatively unaffected by evaluation discordances. According to Goffman (1961, p. 56), "it should be apparent that persons
will differ greatly in their capacity to sustain tension without exhibiting it and without flooding out. This difference in degree of poise is to be accounted for at least in part by differences in group affiliation: middle-class American four-year-olds will sometimes blush and wriggle away when they are merely looked at, whereas Victorian *grandes dames* are reported to have been able to maintain poise under quite disastrous conditions." The "very high" designation is meant to capture any such *grandes dames* effect.

Low, high, and very high groupings were assigned using natural cut points in the media mention data.\(^4\) The low category comprises approximately 30-50\% of the data, depending on the panel year. Although the number of media mentions among golfers in this group fluctuates by year, they range from a minimum of 0 to a maximum of 52 in the final year of the panel. The high category includes the next 40 to 60\% of observations, again depending on the panel year, with media mentions in this group ranging from 53 to 210 in the final year of the panel. The top 10\% of the data, according to the number of media mentions, constitutes the "very high" category and ranges from 210 to 1000+ mentions in the final panel year. By restricting the sample set to each category, I am able to run three simultaneous linear models that allow me to investigate differences in the effect of proximate feedback on risk taking by group. According to the hypothesis, I expect that within-actor variation in risk taking at time \(t+1\) will be greater, on average, for those golfers who at time \(t\) exhibit discordance between their proximate and distant feedback. That is, golfers who receive positive feedback by one audience but negative feedback from the other will react to the discordance by adopting more risky playing strategies. The resulting predictive models take the form:

\(^4\) Cut points were determined in two ways. First, I looked for visually identifiable changes in the first-order differential slopes of the distribution of distant feedback for each panel year. Next, I matched the approximate coordinates of the major slope changes to percentiles. This was important to account for an overall upward trend in the number of media mentions of the median golfer in the panel (27 in 2002 to 52 in 2005). A long right-hand tail describes the data beyond the 90\textsuperscript{th} percentile in each of the panel years, supporting the designation of "very high" for these data points.
\[ risk_{i,t+1} = \rho P_{i,t} + \mu_t + \tau_{i,t+1} + \epsilon_{i,t+1} \]

where individual \( i \)'s average level of risk, \( risk_{i,t+1} \), or attempts at going for the green, over the course of a season is regressed on his proximate feedback, \( P_{i,t} \), in the prior season. \( D_{i,t} \) corresponds the distant feedback-based groups described above. One of the additional advantages of this approach is that mean differences in risk taking by each of the three groups are accounted for by the estimated constant. If the evaluation discordance hypothesis holds, the coefficient on \( P_{i,t} \) should differ across the three regression equations. Of the golfers with low distant feedback \( (D_{i,t} = 1) \), risk should be associated with those who have elevated proximate feedback \( (\rho > 0) \). For golfers with high distant feedback \( (D_{i,t} = 2) \), by contrast, risk should be higher among those with low proximate feedback \( (\rho < 0) \). In short, for the evaluation discordance hypothesis to be supported, golfers with low (high) distant feedback should take more risk when their proximate feedback is high (low). Finally, if the \( grandes dames \) hypothesis holds, the effect of proximate feedback should be negligible among the \( D_{i,t} = 3 \) group.

In each of the models I adjust for all forms of temporal heterogeneity by including a fixed effect for each year in the panel, \( \tau_{i,t+1} \). The use of a fixed season effect absorbs any and all year-level changes that may impact the amount of risk taken by Tour golfers, most notably the introduction of new technology that allows golfers to hit the ball further and more accurately, on average, than in prior seasons. Furthermore, I add a player fixed effect term, \( \mu_t \), to adjust for all time-invariant athlete-specific characteristics—e.g., ability, risk tolerance, etc.—that may impact an individual’s propensity to take more or less risk. Correlations and descriptive statistics for all variables are shown in Table 1.
As I alluded to when discussing the construction of the risk variable, all individual-level predictors are collapsed to span the course of single seasons. For instance, despite measuring proximate feedback on a weekly basis, the variable that enters the models is the season average of a golfer’s proximate feedback. There are two primary reasons for this. First, meaningful changes in golfers’ strategies—specifically the amount of risk they take—are more likely to occur between seasons rather than within seasons. Week-to-week variations are usually attributable to course conditions and the competitive field of a given tournament. Because problemistic search is assumed to be a deliberate response to evaluation discordance—and changes in risk taking therefore a strategic behavior—week-to-week or even month-to-month variation are unlikely result from evaluation discordances. A second advantage is about identification. Due to the panel nature of the data, each of the models are specified such that discordant evaluations in year $t$ can only have an effect on a player’s risk taking in year $t+1$. If a shorter lag structure were used, it would be more difficult to rule out the possibility that players’ proximate feedback affects their distant feedback, or media coverage, at heterogeneous rates. For example, whereas changes in Tiger Woods’ standing among golfers may be immediately reflected by the media, similar changes for Stephen Ames’ may be covered by the media after a somewhat longer delay. For even less prestigious golfers, this delay might be even longer.

**Results**

Initial results are shown in Table 2. Models 1 and 2 include all data and reveal a strong negative effect of proximate feedback on risk taking, controlling for age. Model 2, which differs only in that it fits the effect of age non-linearly, highlights the magnitude of the association between proximate feedback and risk: all else held constant, the effect of proximate feedback on risk taking for a golfer at the 75th percentile (0.935) is approximately twice that of a golfer at the 25th percentile (0.184). The effect of age on risk taking is not in fact curvilinear but asymptotic—the measured inflection point extends beyond the range of the data—and indicates that age has a positive effect on
a golfer’s level of risk, but that the association is positive at a decreasing rate. One explanation for this age effect is simple selection. Older golfers who maintain their strength continue playing PGA events whereas those who see their strength decline opt for earlier transition to the senior Champion’s Tour.

Distant feedback enters the set of models reported on the right hand side of Table 2. The parsing strategy indicates a difference in the effect of proximate feedback across the three groups. Whereas the effect of proximate feedback is positive for those with low distant feedback (0.196, t = 2.51), the coefficient estimate changes sign among those golfers having high distant feedback (-0.072, t = -2.48). In other words, when a golfer’s evaluation discordance is high during a given season—meaning he has either low distant and high proximate feedback, or high distant and low proximate feedback—he is more likely to adopt a riskier playing strategy in the subsequent season.

Among those golfers with the lowest distant feedback, a one standard deviation increase in their proximate feedback corresponds to more than a half standard deviation increase in risk. Over the course of a single tournament, this change amounts to a 14% increase in the number of times an average golfer goes for the green.5 For those with high distant feedback, the effect of a one standard deviation increase in proximate feedback amounts to a 5% reduction in a golfer’s attempts at the green. Finally, the grandes dames effect is supported in the right-most set of results. This last model, however, should be taken as suggestive only: there are simply fewer opportunities for significant evaluation discordances among this group of prominent athletes. Of the 94 data points in this group, only 29 players have proximate feedback that is less than the overall mean for all golfers. Furthermore, I expect that many of these data points correspond to once great players who, over the course of the panel, were transitioning to the PGA seniors’ tour. Consequently, their standing solely among peers should not keep pace with their level of celebrity in the sporting media. A quick scan of

5 The average golfer in this category goes for the green 4.88 times per tournament (1.218 attempts x 4 rounds). A one standard deviation increase in golfers' proximate evaluations amounts to .704 additional attempts over 4 rounds, or 5.584 total attempts.
the names in this grouping confirms this intuition and includes such past greats as Tom Watson, Tom Kite, Greg Norman, and Nick Faldo.

To assess the robustness of the findings in Table 1, Table 2 presents a series of additional regressions that account for possible time-varying changes in players' ability. Despite the use of fixed effects in the first set of models—which control for such things as players' ability and propensities to take risks—it is still conceivable that changes in the amount of risk employed by an individual are impacted by changes in a player's abilities in other facets of the game. For instance, if, through diligent training and practice, a golfer becomes particularly accurate in their approach shots from one season to the next—an approach shot is typically the second shot on a par four or third shot on a par five hole—he may limit the amount of risk he takes from the tee. The same may go for improvements in a player’s putting and the opposite should hold if a player spends the off-season devoting himself to strength training, which should allow him to go for the green more often. To account for possible time varying effects, I include six additional variables, one for the current year and one for the past year, for "putts per round," "greens hit," and "driving distance." The first and last variables are self-explanatory. "Greens hit" is computed by the PGA as the number of approach shots—i.e., shots into the green—that land and remain on the green.

As a final control, I enter the average level of tournament prestige a golfer plays in during the course of a given season, measured by the size of the tournament purse. This additional variable serves three functions simultaneously. First, it adds yet another control for a player’s quality. Better players will be invited to play in more prestigious, and thus more lucrative, tournaments. Second, the variable provides a crude but reasonable proxy for the amount of endorsements received by an individual golfer. Although there is significant variance in the number of sponsors supporting a player and the amount of endorsements received, playing in more prestigious events increases a player’s television exposure and should, therefore, have a significant impact on endorsements. Third, while I have no predictions about the relationship between tournament prestige and risk, it is feasible
that players might adopt more or less risky playing strategies when they enter more visible and lucrative tournaments.

Consistent with expectations, season-to-season improvements in a player’s average driving distance and putting have positive and negative effects on his propensity to take risks, respectively. Interestingly, golfers who become more adept in the accuracy of their approach shots take more, not less, risk. Perhaps this a reflection of becoming more accurate in general, including shots from the tee. Purse size has a strong and negative effect on risk taking. On average, high stakes begets conservative play. Importantly, the effects of proximate feedback across the four models shown are consistent with the earlier models.

Discussion

My aim in this chapter has been to offer a theoretical reconfiguration of an old sociological intuition. Feelings of strain and anxiety, as well as behaviors meant to alleviate those feelings, result from conflicting evaluations of an individual’s performance in a single role. The results of my analysis of PGA Tour players support this proposition, indicating not only that professional golfers respond to multiple streams of feedback, but also that the relationship between such streams is an important predictor of subsequent behavior. When there was a disconnect between the feedback players received from sports media and their standing among peers in one season, players exhibited additional risk taking behavior in the following season.

Naturally, this chapter is but one empirical investigation and replication of its results in alternative contexts would do much to strengthen the evaluation discordance hypothesis. In addition to replication, further studies might also consider the possibility that individuals apply greater weight to one audience's feedback over another. I addressed the possibility of differential weighting in this chapter only in theory—by stressing that for people to feel strain as a result of misaligned evaluations, they must consider each of the audiences to be relevant at the point in time when
evaluations are made or feedback is given. In other words, I assumed that golfers consider proximate and distant feedback to be equally important. Future research might relax this assumption in order to attend to questions about people's motivations—a self enhancement motivation, for instance, might lead people to ignore feedback that is highly relevant but negative—in addition to their need for evaluative consistency.

**Organizational Implications and Additional Research**

Both the theory and empirical context in this chapter are fundamentally about people. It is people in my setting who receive conflicting evaluations of their performance and people who respond to this conflict. This volume, however, intends to bring together research on organizational science. Consequently, I am obliged to extend the findings here to consider possible implications of my results on organizational processes. One of the key contributions of this chapter, I believe, is to offer new material for rethinking the nature and consequences of multiple audiences—and specifically feedback from multiple audiences—in social and economic contexts. Unlike prior organizational research on performance feedback (e.g. Cyert and March, 1963; Greve 2003) wherein actors are assumed to determine a set of relevant social referents against which to compare some meaningful outcome (like performance), my findings indicate that multiple sources of feedback may themselves become the object of comparison.

Certainly this is not the first study to take multiple sources of feedback into consideration. In management research, for instance, Kacperczyk, Beckman, and Moliterno (2011) recently examined the response of financial managers to both internal (to the firm) and external social comparisons. The authors find that "the internal social comparison is more salient to decision makers than external social comparison for motivating change." Similarly, in finance, it is common for asset pricing models to include not one but several known reference points, or benchmarks, against which to evaluate the return of a given asset. Whereas both of these approaches treat each reference group or benchmark as unique, however, my results suggest that the relationship among different sources of
feedback may also be important. Managers and investors alike should always desire their performance to outdo as many of their evaluative benchmarks as possible. It is also feasible, however, that additional satisfaction is gained by outperforming each of the benchmarks in approximately the same way—i.e., when there is consistency across the multiple sources of feedback.

When it comes to predicting risk-taking behaviors, specifically, existing organizational research is equivocal, sometimes showing that outperforming a benchmark or reference group leads to more risk, sometimes showing that underperforming the benchmark or reference group leads to more risk, and sometimes showing no effect at all (see Kacperczyk et al. 2011). My findings highlight the fact that social actors orient themselves to more than one source of evaluative feedback. Accordingly, existing research on performance feedback in organizational contexts may benefit from an expanded scope. Risk may not systematically result from outperforming or underperforming a given reference set, but rather be the consequence of multiple and contradicting feedback.
References


