



Implicit Motive Profile Analysis: An If-Then Contingency Approach to the Picture-Story Exercise

Oliver C. Schultheiss* and Maria Schultheiss

Friedrich-Alexander University

Abstract

Picture-story exercise (PSE) measures of implicit motives provide objective, stable, and valid scores. However, PSE scores are also characterized by substantial variability from one picture to the next, resulting in low internal consistency estimates. We argue that this variability is a critical source of the PSE's validity because it reveals stable if-then contingencies between the situational contexts represented by the picture cues and individuals' motivational responses to them. We also present the correspondence hypothesis, according to which the situational cues in the PSE are representative of specific real-life situations and a person's imaginative-story responses to the PSE cues therefore also reflect his or her real-life responses to such situations. We review corroborating research and discuss the implications of the if-then contingency approach for assessment, prediction, and for the measurement of motivational congruence.

Introduction

Implicit motives are enduring dispositions to experience certain classes of incentives as rewarding (Schultheiss, 2008). There are four motives that are most frequently studied. The need for achievement (*n* Achievement; McClelland, Atkinson, Clark, & Lowell, 1953) is a capacity for enjoying the mastery of challenging tasks. People high in *n* Achievement show high task performance, particularly if they can set goals and perform the task on their own and if they get feedback. Achievement-motivated people also excel in business. The need for power (*n* Power; Veroff, 1957; Winter, 1973) is a capacity for deriving pleasure from having an impact on others. Power-motivated individuals seek impact by, for instance, persuading others and drawing others' attention to themselves through extreme risk taking or by owning prestigious objects. They excel in leadership and management positions. The need for affiliation (*n* Affiliation; Atkinson, Heyns, & Veroff, 1958) is a capacity for reveling in friendly, harmonious relationships with others. Individuals high in *n* Affiliation strive to be included by like-minded people and are rather sensitive to rejection. The need for intimacy (*n* Intimacy; McAdams, 1980) is a capacity for cherishing deep, meaningful relationships. Like affiliation-motivated individuals, intimacy-motivated people like to be with others but in the context of deep, personal relationships. All four motives are implicit in the sense that they operate nonconsciously while influencing patterns of thought, feeling, physiology, and behavior (McClelland, Koestner, & Weinberger, 1989).

Implicit motives are typically assessed with the picture-story exercise (PSE; McClelland et al., 1989), a descendant of Morgan and Murray's (1935) Thematic Apperception Test (for reviews of PSE motive measures, see Atkinson, 1958a; McClelland, 1987; Smith, 1992; Schultheiss & Brunstein, 2010). On the PSE, test takers are asked to write imaginative stories about four to eight pictures showing people in various situations. Some of the most frequently used PSE pictures show a captain talking to a passenger, two women working in a laboratory, a boxer, two people on a bench next to a river, a man and a woman in a

nightclub, and two trapeze artists (Pang & Schultheiss, 2005; for picture reproductions, see Ramsay & Pang, forthcoming, and Schultheiss, 2013). Coders who have achieved at least 85% coding agreement with experts later code the stories for motivational themes based on established coding systems for motive assessment. The sum of all motivational themes across a test taker's stories, adjusted for total protocol length, represents his or her motive score (see Schultheiss, 2013, for an illustrative example of coding a story for motivational themes).

PSE coding systems for the assessment of implicit motives were developed by arousing motivational states in one group of participants and comparing their stories with stories that non-aroused participants in a control group wrote about the same pictures. Thematic differences between stories written under arousal versus control conditions were then distilled into specific coding rules. For instance, in contrast to control-group participants, individuals whose need for power had been aroused by inspirational speeches or being in a position of power over another person wrote stories more saturated with story characters having a forceful impact on others, stories with themes of manipulating, persuading, or impressing others, and stories in which strong emotions in others were elicited (Winter, 1991). The coding system for *n* Power is based on these indicators of aroused power motivation (Winter, 1973, 1994).

The experimental-arousal approach was based on the notion that a measure should be sensitive to experimental variations in the phenomenon it aims to capture (McClelland, 1958, 1987, pp. 200–202), akin to validating a thermometer by showing that it gives a low reading when immersed into freezing water and a high reading when immersed into boiling water. Recently, Borsboom, Mellenbergh, and van Heerden (2004) argued that this type of *causal validation* represents the most direct and critical evidence for a measure's validity.

In the same manner that a thermometer can be used to determine the temperature of new substances, coding motive themes in picture stories has been used in many studies to assess dispositional differences in motive strength – that is, to gauge individuals' typical motivational “temperature”. This approach yielded impressive results in terms of predicting a broad range of outcomes, from physiological processes to societal changes (see McClelland, 1987; Schultheiss, 2008, and Spangler, 1992, for summaries and meta-analyses). However, the reliability of the PSE, particularly its internal consistency, has remained a contentious issue (e.g., Entwisle, 1972; Lilienfeld, Wood, & Garb, 2000). In the present paper, we will argue that such concerns are primarily due to ill-conceived theorizing and a lack of attention to clearly observable phenomena inherent in the picture-story approach to motive measurement. We will present an emerging new approach to understanding the PSE that is based on a careful examination of the interplay between the picture cues used to elicit stories and the thematic response categories that constitute the coding systems. We will also sketch out novel diagnostic possibilities that result from this approach, with a partial focus on the issue of congruence between implicit motives and the motivational needs that people consciously attribute to themselves.

The Reliability Issue

PSE motive measures show four characteristic properties (Schultheiss & Pang, 2007): (i) After sufficient training with the coding systems, coders can reach high inter-rater reliability, with coefficients frequently exceeding 85% agreement; (ii) PSE motive measures have substantial retest stability; (iii) PSE motive measures have low internal consistency, with scores fluctuating substantially from one picture story to the next for a given participant; and (iv) in the presence of proper motivational incentives, motive dispositions assessed with the PSE predict relevant and substantial outcomes such as instrumental learning, social behavior,

and physiological responses. Property 3 – low internal consistency – has vexed both proponents and detractors of the PSE approach to motive assessment. The detractors used the PSE's low internal consistency to question the validity of the PSE (e.g., Entwisle, 1972; Lilienfeld et al., 2000), ignoring that the PSE has well-documented predictive validity and is one of the very few personality measures with demonstrated causal validity (see Borsboom et al., 2004). The proponents either claimed that internal consistency is not a relevant reliability criterion for the PSE (e.g., McClelland, 1980), an argument that is defensible according to modern psychometric theory (e.g., Borsboom, 2005), or they argued that low internal consistency is a result of insufficient test length (e.g., Blankenship & Zoota, 1998; Heckhausen, 1963). Few have recognized that low inter-story reliability is a potentially valuable phenomenon that should be taken seriously and understood in its own right.

A well-known exception to this rule is Atkinson's dynamics-of-action (DOA) theory account of low internal consistency (Atkinson, 1981; Atkinson & Birch, 1970). Atkinson argued that observable variations between high and low scores from one picture story to the next are the result of the intrinsic dynamics of motivation. When a motive is aroused by a given picture cue, it is expressed in a person's story. However, the expression also reduces need strength temporarily, making the expression of that motive in subsequent stories less likely. If and when the motive shows up again in one of the next stories depends on the following: The stronger the motive, the quicker it can be aroused again by subsequently presented picture cues, resulting in a more rapid oscillation between high and low scores in a person's PSE protocol. This theory suggests that a person with a strong motive should have less internally consistent scores due to the rapid waxing and waning of motivational states. This prediction was indirectly supported by computer simulations (e.g., Atkinson, Bongort, & Price, 1977) and by an empirical study in which PSEs with lower internal consistency had better predictive validity than PSEs with higher internal consistency (Reuman, 1982). However, when the central prediction of DOA, namely that the expression of a motive should reduce its expression on a subsequent story, was directly tested in a study with a large sample, the results failed to support the model (Tuerlinckx, De Boeck, & Lens, 2002).

A more mundane reason for low internal consistency pertains to the use of heterogeneous pictures. Schultheiss and Brunstein (2001) and Pang and Schultheiss (2005) observed marked differences in the number of coded themes that different pictures elicit for a given motive. In response to *couple by river*, one frequently used picture (see Schultheiss, 2013), participants respond on average with two codeable instances of affiliation but with zero instances of achievement. By contrast, the picture *women in laboratory* tends to elicit one achievement theme and zero affiliation themes (see Ramsay & Pang, forthcoming). Thus, *couple by river* has a high "pull" for affiliation themes and a low pull for achievement themes, whereas the reverse is true of *women in laboratory*. Schultheiss and Pang (2007) have suggested that if a PSE consists of such diverse pictures, it may have broad validity for many different predictive contexts because it samples individuals' motivational responses to very different situations. However, precisely because the situations depicted are so different in terms of their pull for a given motive, internal consistency will be low. Reducing the cue bandwidth of the PSE by using thematically more similar pictures will lead to more consistent responses but will also narrow the PSE's predictive validity to only very specific situations at the expense of others (see below for further considerations of this type).

Despite the low internal consistency of motive scores on broadband PSEs like the one described by Schultheiss and Brunstein (2001) and Pang and Schultheiss (2005), most researchers aggregate scores for a given motive across pictures into a sum score. McClelland (1980) justified this practice by arguing that motive themes in response to different pictures

represent alternative manifestations of the same underlying motivational need, much as one hungry person would eat pasta, but not pie, while another hungry person would prefer pie over pasta.

Taking an opposing point of view, Fleming (1982; see also Fleming & Horner, 1992; Lundy, 1985) has argued that because individuals' responses to different pictures do not strongly covary, they should *not* be aggregated into a sum score. Referring to some of her own findings, Fleming (1982) stated:

“[...] motivation research relies on the indiscriminate summation of cue scores, but here this procedure resulted in dramatic losses of predictive power, from 49 percent to 12 percent of the variance when compared to the results from a regression analysis [...]. The clear implication is that the magnitude of the findings in the literature could be severe underestimates of true relationships”.(p .85)

To illustrate Fleming's (1982) point, we have reanalyzed data from a study originally published by Schultheiss and Brunstein (2002). In this study, participants had to convince another participant of their point of view on the defensibility of animal experimentation. Their performances were videotaped and later rated by judges for perceived persuasiveness. Schultheiss and Brunstein (2002) found that a combination of high *n* Power and high activity inhibition, a moderator of motive expression, was predictive of judged persuasiveness in participants in a motive-arousing goal-imagery condition but not in a control condition. Focusing on participants in the critical goal-imagery condition for present purposes, we simultaneously regressed disaggregated *n* Power scores from the six-picture cues used by Schultheiss and Brunstein (2002) on judged persuasiveness. We find that the variance explained by this multiple regression is a magnitude larger than that explained by the total sum score, even when adjusting for number of predictors involved (see Table 1). This finding clearly corroborates Fleming's (1982) earlier observations.

On the basis of such results, we would like to suggest that there is predictive validity to picture-specific responses that is lost by the calculation of sum scores. The next section provides a theoretical context for this suggestion.

Table 1. Predictive validity of *n* Power picture scores (simultaneous regression) versus a *n* Power sum score for judged persuasiveness in the goal-imagery condition of Schultheiss and Brunstein's (2002) study ($n = 34$).

Picture	<i>B</i>	<i>p</i>
Ship captain	0.61	.001
Trapeze artists	0.19	.29
Couple by river	0.00	.99
Nightclub scene	−0.03	.88
Architect at desk	−0.11	.52
Women in laboratory	−0.52	.007
	$R^2 = .407$ (adjusted $R^2 = .275$), $p = .02$	
<i>n</i> Power sum score	0.18	.31
	$R^2 = .032$, $p = .31$	

The PSE as a Measure of Motivational If-Then Contingencies

Before he developed DOA theory, Atkinson (1958b) provided a social-learning theory explanation to account for picture-story variability. He argued that a person's response to each picture is a function of the person's motive strength and specific expectations he or she associates with each cue (see Figure 2, p. 608, in Atkinson, 1958b). He used the following example to make his case: Imagine two individuals who, due to early childhood experiences, have both developed a strong achievement motive. One quits school early, gets vocational training, and is socialized into a blue-collar job. The other gets a high school degree, attends college, and then goes into business. Both are administered two PSE picture cues. One shows a man working in a shop (i.e., a blue-collar achievement situation), whereas the other shows a business office scene (i.e., a white-collar achievement situation). Atkinson (1958b) wrote:

“For the first young man, achievement-related expectancies should be relatively strong in response to the blue-collar picture and relatively weak in response to the white-collar picture. Just the reverse is true of the second young man. Given equally strong motives to achieve, the frequency of achievement-related imaginative responses should be greater in response to the blue-collar picture for the first young man and greater in response to the white-collar picture for the second young man. But their total scores from the two stories should be very comparable”. (pp. 609–610)

Unlike in his later DOA theory, Atkinson (1958b) here argues that the two individuals he describes show different responses to the same pictures due to their different learning histories. These, he contends, have shaped the degree to which the two individuals have come to associate different situations with opportunities for achievement. Similar arguments about the link between a person's learning history and his or her responses to specific picture stimuli were also advanced by Auld (1954) and Goss and Brownell (1957).

Atkinson's (1958b) account of PSE picture score variability presaged modern theories of personality, most notably Mischel and Shoda's (1995) cognitive-affective system theory. Like Atkinson (1958b), these authors argue that people stably construe some situations, but not others, as opportunities to express a specific disposition. According to their theory, stability in personality results from stable profiles of if-then contingencies between specific situational stimuli and behavioral responses.

Schultheiss, Lienesch, and Schach (2008) applied these ideas to the PSE and hypothesized that it represents a measure of if-then contingencies between specific situational contexts, as represented in the picture cues (i.e., the *ifs* of behavior), and a person's likelihood of responding with motivational themes in her or his stories (i.e., the *thens* of behavior). More specifically, Schultheiss et al. (2008) hypothesized that although an individual's motivational responses may vary between picture cues, *the if-then response profile resulting from this variation should be stable from one testing occasion to the next for a given individual*.

To test this idea, Schultheiss et al. (2008) administered the same eight-picture PSE twice, separated by 2 weeks, to 90 US research participants. Because the same set of pictures was administered twice, participants were told at the second session that they were free to write similar stories as the first time or to write different ones. This instruction took into account that the standard PSE instruction to write *imaginative* stories may induce some test takers to intentionally make their stories as dissimilar as possible on the second administration, a test set that would artificially reduce retest correlations (see Lundy, 1985; Winter & Stewart, 1977). All PSE stories were coded for motive imagery by two trained coders who were blind with regard to the within-subjects test session factor and who achieved satisfactory to high interscorer reliability (Pearson's $r_s = .70$ to $.86$). The resulting sum motive scores showed

substantial between-session retest reliability, with coefficients ranging from .37 to .61. However, as in previous research, within-session internal consistency coefficients were low, with a range from $-.02$ to $.43$.

To test for PSE if-then profile stability, Schultheiss et al. (2008) first removed the influence of story length and the average pull of each picture for a given motive and then calculated, for each motive separately, correlation coefficients for a 2 (testing occasion) \times 8 (picture story) design. Such procedures retain what is *distinctive* to each story response and can be considered a “profile” (see Furr, 2009). Significant picture–response retest correlations ranged from .21 to .40. This range is similar to the behavioral profile stability coefficients reported by Shoda, Mischel, and Wright (1994). As the case example in Figure 1 illustrates, these stability coefficients suggest that each participant’s profile of responding to the eight-picture cues remained moderately stable from one testing occasion to the next. Such correlations can explain why summed retest stability can be high when internal consistency (within a session, but across pictures) is low (Schultheiss et al., 2008).

Busch and Hofer (2012) replicated these findings with a shorter PSE (five picture cues), a longer retest interval (18 months), and in larger samples from two different cultures (129 Germans and 122 Cameroonians). Except for *n* Power scores in the Cameroonian sample ($r = .06$), people were consistent in their if-then motive profiles over time, with correlations ranging from .13 to .20. The lower level of profile stabilities compared to those reported by Schultheiss et al. (2008) is likely due to the much longer retest interval, the shorter PSE employed, and protocol lengths that were quite different between sessions. Overall, however, the Busch and Hofer (2012) results corroborate those obtained by Schultheiss et al. (2008). That is, both studies provide compelling support for the idea that the PSE captures stable if-then contingencies between specific situational cues and motivational responses, with further potential implications for the predictive validity of the PSE method.

Implications for the PSE’s Validity

The observation that the PSE measures stable if-then contingencies between situational contexts and motivational responses has consequences for our understanding and diagnostic use of PSE motive scores. The first consequence is that motive scores summed across all

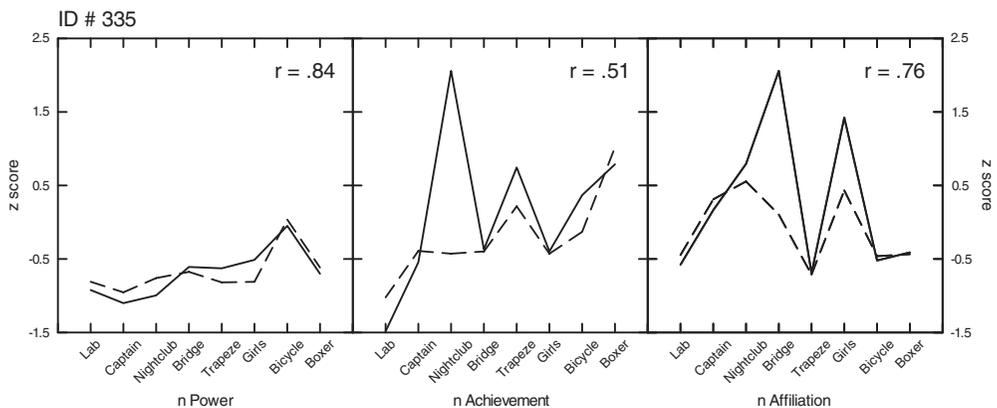


Figure 1. Ipsative stability of a participant from Schultheiss, Liening, and Schad’s (2008) study, calculated as the within-person correlation between eight-picture scores on the PSE at time 1 (solid line) and at time 2 (striped line, 2 weeks later).

picture stories conflate two properties. One is the *extensity* of a motive – that is, *to how many different situational contexts (i.e., picture cues) a person responds motivationally* (see McClelland et al., 1953, pp. 213–214). For instance, a person who responds to all six-picture cues of Pang and Schultheiss's (2005) PSE with power themes has a power motive that is sensitive to more situational contexts than a person who responds only to three cues with power themes. However, the issue is complicated a bit by the fact that sum scores capture not only a motive's extensity, but also its *intensity* – that is, *how strongly a person responds to each picture* (McClelland et al., 1953, p. 214). For instance, a power motive sum score of six can be obtained by a person injecting exactly one codeable power theme into each of the six stories he or she has written or by responding to only one picture with six instances of power motive themes.

Consistent with this analysis, we find in a data set with 161 German participants that extensity (the number of stories with a codeable theme) and intensity (the maximum number of codeable themes for one story) both predict motive sum scores in a multiple regression, all $Bs > .39$, all $ps < .0001$. Extensity and intensity parameters also interact to predict sum scores, all $Bs > .28$, all $ps < .05$, such that sum scores are the highest when both extensity and intensity are high. Sum scores thus conflate extensity and intensity while obscuring meaningful if-then variance. This implies that it may be problematic to create sum scores and that one should take the variance between picture cues seriously and treat it as diagnostically useful information, as Fleming (1982) recommended. McClelland et al. (1953) made similar points:

“[...] one could imagine constructing a motivational profile for a given individual which would show the intensity of his achievement motive in a variety of situations. Our present *n* Achievement score would then turn out to be a composite in which extensity and intensity of motivation are confounded. This is an area of research which clearly needs further exploration”. (p. 214)

Consider Figure 2, in which the power imagery of two individuals is plotted by story. Summed power scores would be quite similar yet would mask important heterogeneity. The individual on the left responds with power motive imagery predominantly to those

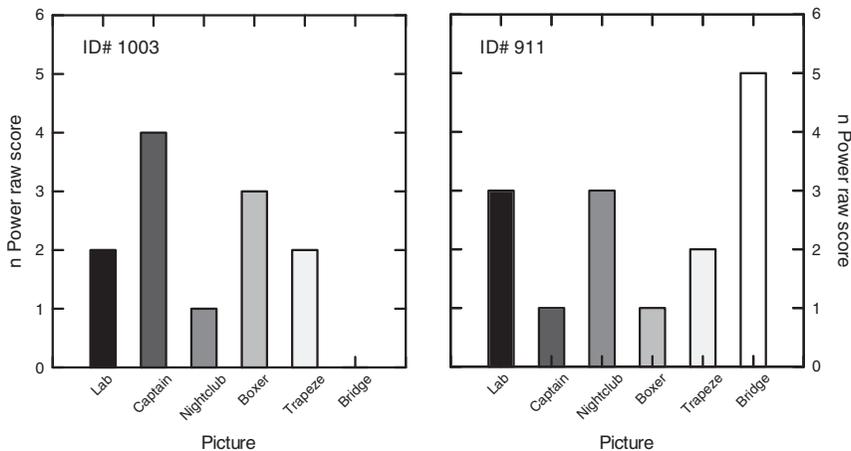


Figure 2. Picture cues are ordered from the highest pull for *n* Power (left, black) to the lowest pull for *n* Power (right, white), based on the sample's average picture scores. Participants 1003 and 911 are roughly similar with regard to total word count (721 and 823, respectively) and *n* Power sum scores (12 and 15, respectively). Participant 1003 tends to write more power-saturated stories in response to more prototypical pictures.

pictures that, averaged across many test takers, have a strong pull for power (e.g., at work, when interacting with an authority figure); by contrast, there is no power imagery for a picture depicting an intimate conversation. The individual on the right, instead, responds with the largest amount of power imagery to this intimacy-related picture! We might expect the latter, but not former, person to exercise power in intimate situations.

Consistent with Mischel and Shoda's (1995) account, we further propose *that there should be a correspondence between individuals' imaginative-writing responses to picture cues in the PSE and their actual responses to similar situations in real life* (see Figure 3). Again consider the two individuals of Figure 2. By the correspondence principle, the occurrence of power-related behaviors in intimate relationships should be particular to the second person but not the first.

Of course, we do not wish to suggest that the correspondence between PSE picture cues and real-life situations is a literal one. For example, we doubt that responses to the captain picture only possess validity in predicting interactions with real-life captains. Instead, we assume that validated PSE pictures can be viewed as prototypical of a class of situations. If so, the captain picture is prototypical of situations in which one has, or is dealing with, formal authority and is trying to influence others through arguments and persuasion. Similarly, the laboratory picture may capture situations in which one works with others, in which work quality is critical, and in which there is an element of supervision and competition, etc.

The correspondence principle has additional implications for scoring and assessment. We argue that in addition to an overall motive-imagery score for each picture story, one can also

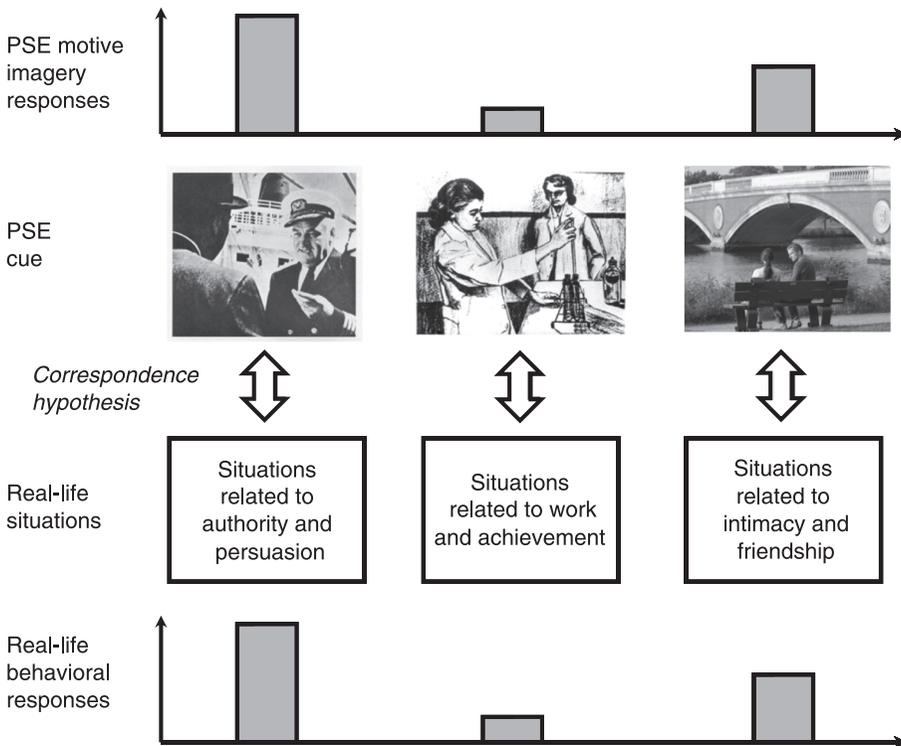


Figure 3. Model of the correspondence between if-then contingencies as expressed in the PSE and if-then contingencies expressed in real-life behavior. Upper half: amount of motive imagery present in an individual's responses to various PSE picture cues. Lower half: behavioral responses indicative of a motive in response to real-life situations corresponding to PSE pictures.

distinguish *specific types of imagery responses* for a given motive across picture stories. In the case of n Power, Winter's (1994) running-text system, which is one of the most frequently used coding systems, differentiates between six response categories that are all clearly distinct but share an inherent concern with having an impact on others. According to Winter (1994), power imagery can be coded in the following instances: (i) strong, forceful actions, such as yelling at someone or aggressing in some other way; (ii) control or regulation, such when someone is spying on someone else or tells others what to do; (iii) arguing with, persuading, or convincing others; (iv) providing unsolicited help and advice to others; (v) trying to impress others or showing a concern with fame prestige; and/or (vi) eliciting strong emotional responses in others, such as startling them, making them laugh, cry, etc.

We suggest that people will meaningfully differ in which of these (Winter, 1994) power themes they typically use. As an illustration, the individual depicted on the left in Figure 4 expresses his or her concern with impact on others primarily with imagery related to strong, forceful actions, control and regulation, and impressing others but shows little inclination to influence others through more prosocial means such as persuading or helping. The individual depicted on the right, however, expresses his or her need for power much more often through such prosocial types of imagery. If the correspondence hypothesis holds for the response side as it does for the cue side of the PSE, then we would expect the former, relative to the latter, individual to express her or his need for power in a more self-centered, competitive, and perhaps even aggressive manner in real life.

Although no studies have been published so far in which the stability of motivational response profiles has been tested in a similar manner as has been done for picture profiles (i.e., Busch & Hofer, 2012; Schultheiss et al., 2008), other evidence provides preliminary support for the proposed extension of the correspondence principle to the response-facet side of imagined and real-life behavior. Schultheiss, Campbell, and McClelland (1999) report that individuals who are predominantly motivated by a self-centered type of power

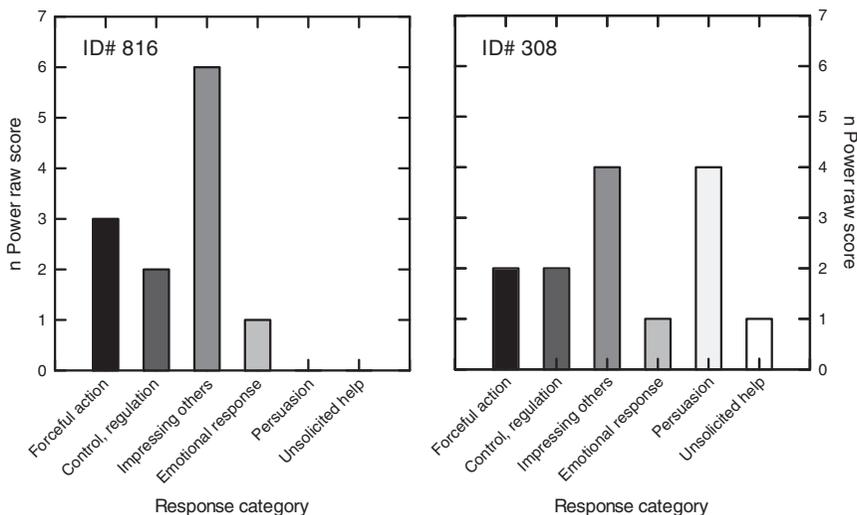


Figure 4. n Power response categories are ordered from least socialized (left, black) to most socialized (right, white). Participants 816 and 308 are roughly similar with regard to total word count (818 and 711, respectively) and total n Power (12 and 14, respectively). Participant 816's stories are predominantly about impressing others and being famous but also about being forceful and controlling. Overall, participant 816 appears to be characterized by a personalized need for power. Participant 308 appears to strike a better balance between personalized and socialized ways of expressing a need for power.

motive (in their PSEs) respond to a competitive encounter with another person with stronger endocrine signs of motivational arousal than individuals who show evidence of a prosocial power motive. Presumably, the competition situation provided more of an incentive for the former than for the latter individuals.

Also consider the results of McAdams and Powers (1981) on *n* Intimacy. These researchers had 43 participants first complete a PSE and then, 5 to 8 weeks later, participate in psychodrama sessions that were videotaped. Both sources of data were scored for several manifestations of *n* Intimacy (e.g., positive affect during interaction, surrender of control to others, expression of personal meaningfulness, etc.). The correlation between the presence of these response categories in the PSE stories and in participants' actual behavior during the group sessions was a whopping .70! At least part of the reason for this excellent correspondence was likely the match of PSE categories to behavioral criteria.

To sum up, we argue that motive scores aggregated across picture stories provide a summary index of the extensity and intensity of a motive while obscuring information that is relevant for diagnostic and predictive purposes. Such a loss of information need not occur. The relevant information is available in differences (a) between individuals' motive scores from one picture story to the next and (b) between individuals' response styles within a given motive. If so, taking this information into account can be useful for (a) predicting, based on the PSE, which types of situations a person will respond to motivationally versus not and (b) which types of responses he or she will be more or less likely to exhibit. Both such points follow from the correspondence hypothesis.

Implications for Motive Assessment and Predictive Validity

The obvious practical implication of our model for the assessment of implicit motives is a reiteration and extension of Fleming's (1982) earlier recommendation: The prediction of relevant outcomes via sum scores should be complemented, and where reasonable and justifiable even replaced, by simultaneous multiple regressions using (a) separate picture scores and (b) separate response category scores for each motive. Such regressions should control for story length, which predicts codeable motive imagery (see Schultheiss & Pang, 2007).

If the correspondence hypothesis is correct, the precise composition of the PSE is critical, too. For studies that aim at capturing how motives play out in a variety of different situations in real life, such as Brunstein, Schultheiss, and Grässmann's (1998) study of the effects of implicit motives and personal goal pursuits on emotional well-being over time, a *broadband PSE* that features a sufficient number of diverse, but highly representative situational cues will provide the best predictive validity. The six-picture PSE used by Brunstein et al. (1998; see Schultheiss & Brunstein, 2001, and Pang & Schultheiss, 2005, for detailed analyses of this PSE) depicted situations related to work (alone, with others, and in a formal hierarchy) and close relationships (in an intimate setting, in a public setting, and in a situation in which one partner's life depends on the other). It thus tapped the fundamental areas of agency and communion (see Bakan, 1966). One could well imagine that had the PSE been too narrow (e.g., all work-related pictures), the well-being outcomes may have remained unpredicted.

For studies that aim at the prediction of motivated behavior in one specific situation, on the other hand, a *focused PSE* modeling the situation may yield higher predictive validity. In this connection, recall the study on persuasive communication by Schultheiss and Brunstein (2002). These researchers, at that time not aware of the arguments presented here, used a broadband PSE for a focal prediction. As the re-analysis in Table 1 shows, the power motive score from the one picture conceptually most similar to the persuasion task – the captain talking to or arguing with another person – provided bull's eye validity, whereas most other

pictures, all of them rather dissimilar to the persuasion-task situation, contributed little or even negatively to the prediction. We speculate that if Schultheiss and Brunstein (2002) had used three or four PSE picture cues suggesting themes of persuasive communication between people instead, their predictive validity at the sum-score level would have been markedly better. Schultheiss et al. (2005), in contrast, successfully employed a focused PSE for the prediction of endocrine responses and implicit learning in the context of a one-on-one competition in the laboratory. They hypothesized that this situation would be an incentive for individuals high in *n* Power and particularly to those whose power motive was tied to aggressive or competitive dominance. Schultheiss and colleagues therefore assembled five-picture PSEs consisting almost exclusively of pictures suggesting aggression and competitive dominance and tested their hypothesis in two studies, one with men and one with women. *n* Power scores derived from these focused PSEs consistently predicted enhanced implicit learning among competition winners and impaired implicit learning among competition losers as well as endocrine responses to the contest outcome (see also Wirth, Welsh, & Schultheiss, 2006, and Stanton & Schultheiss, 2007, for related findings and replications with focused PSEs).

What such considerations boil down to is that the PSE should not be used as a fixed black box that yields a unitary motive score that miraculously predicts (or fails to predict) some outcome. Rather the PSE should be used by thinking, reflecting researchers who look beyond the sum score, who take specific eliciting cues and imagery responses into account, who try to match their PSEs to the situations in which they want to predict behavior, and who thus use the PSE as a tool box that can be flexibly assembled to fit one's prediction goals.

Implications for the Measurement of Motivational Congruence

The if-then contingency account of PSE motive measurement also possesses implications for modeling the degree of congruence between individuals' implicit and self-attributed, explicit motivational needs. Motivational congruence – that is, a good match between the strength of one's implicit and explicit motives – is considered by many to be an important prerequisite for well-being, life adjustment, and health (see Brunstein, 2010; Hofer & Busch, 2013; Thrash, Maruskin, & Martin, 2012).

Traditionally, the degree of congruence between individuals' implicit and explicit motives has been assessed through difference scores or statistical interaction terms. These and other, more recent methods (e.g., Kazén & Kuhl, 2011) all suffer from the same basic problem, namely the incommensurability between implicit and explicit motive measures (see Schultheiss, Yankova, Dirlikov, & Schad, 2009). How do we know, for instance, whether a PSE *n* Power sum score of six is equivalent to a Personality Research Form (Jackson, 1984) dominance scale sum score of 13? So far, no method of measuring congruence has solved this fundamental problem.

The approach we endorse here can resolve such comparability problems by looking not at motive sum scores but by shifting the focus to *profiles* of implicit motivational responses to specific picture cues and examining the fit of this profile with explicit, self-attributed responses to the same cues. This has become possible with the development of the PSE-Q (Schultheiss et al., 2009), an explicit motive measure that is commensurable with the PSE because it uses the same picture cues and its items represent the response coding categories of Winter's (1994) popular system for the coding of power, achievement, and affiliation imagery. In fact, each of Winter's 15 coding categories (six for power, five for achievement, and four for affiliation) is represented by a self-descriptive item, such as "In this situation,

I would try to impress other people” for the n Power coding category *impressing others*. These 15 items are presented with each PSE picture, and picture-specific scores can be derived by summing the number of endorsed items for each motive domain.

As illustrated in Figure 5, PSE-Q item scores for a given motive can vary across pictures as much as PSE imagery scores. In this diagnostic situation, the critical question is no longer “Is the implicit score as high as the explicit score?” but turns into “How well do a person’s self-attributed responses across picture cues track variations in the person’s implicit motive scores?” The latter question can be answered by looking at the degree to which implicit and explicit responses to the pictures covary, as defined in terms of a within-subject (i.e., ipsative) correlation coefficient. According to this congruence criterion, the individual on the left-hand side of Figure 5 had a very accurate, or congruent, view of his implicit motivational responses to the situations depicted in the PSE, as signified by a high positive ipsative correlation. In contrast, the individual on the right-hand side of Figure 5 explicitly endorsed responses to the picture cues that were largely at odds with her implicit motivational responses, as signified by a negative ipsative correlation coefficient. Note that this way of measuring motivational congruence depends on within-subject covariance, not sum scores among the sample as a whole.

Schultheiss, Patalakh, Rawolle, Liening, and MacInnes (2011; Study 3) conducted the first published study on the validity of ipsative congruence scores. They assessed implicit and explicit motives with the PSE and the PSE-Q, respectively, in a sample of 99 research participants. For each motive, two types of congruence scores were derived: (a) a traditional measure of absolute differences scores, based on individuals’ sample-standardized motive sum scores and (b) ipsative correlation scores in the manner described above. The results can be summarized as follows: First, ipsative congruence

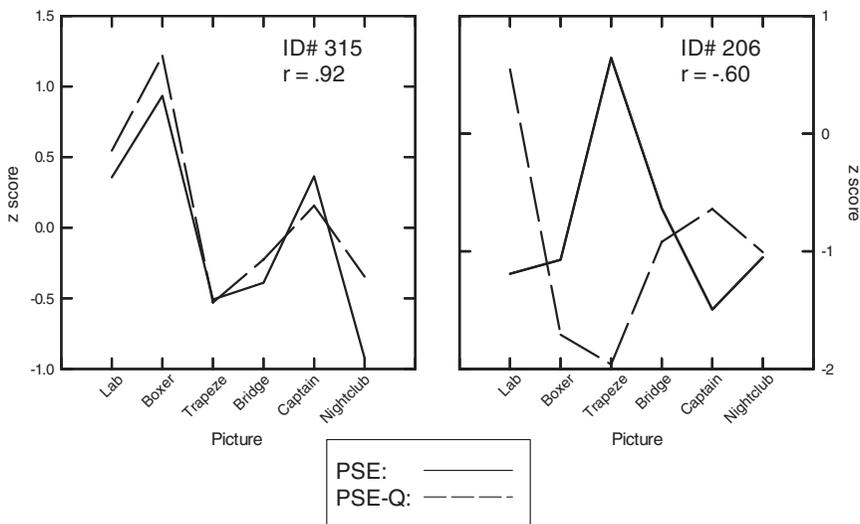


Figure 5. Congruence between PSE and PSE-Q picture profiles for n Power. On the PSE, picture-story word count was partialled from the n Power score, and residuals were converted to z scores. On the PSE-Q, scores were standardized for each picture. Both procedures result in scores that are unrelated to a given picture’s normative pull for motive imagery (PSE) or item endorsements (PSE-Q) or to writing fluency (PSE; see Schultheiss et al., 2011, for further details). Participant 315 shows high congruence between his response profile on the PSE-Q and the amount of power imagery he expressed in response to different PSE pictures, as indicated by a high positive correlation between both profiles. In contrast, participant 206 shows low congruence between her responses on the PSE-Q and the content of her PSE stories.

scores differed significantly from zero, suggesting that participants were able to correctly identify which situations they responded to at the implicit motivational level. However, the average ipsative correlation was $r = .120$, consistent in sign and approximate size with Spangler's (1992) meta-analysis. This suggests that for people in general the level of explicit insight into their implicit motivational responses is very limited, even when the focus is shifted from global score correlations to ipsative profile correlations.

Second, ipsative correlations were not predictive of an absolute difference score measure of congruence, underscoring our previous argument that within-subjects congruence represents something fundamentally different than what between-subjects congruence represents. And third, Schultheiss et al. (2011) showed that higher ipsative congruence scores were associated with better verbalizing ability (termed *referential competence*), a finding that is consistent with the idea that accurate self-knowledge critically depends on language for a person to become cognizant of his or her inner world of experience (see Schultheiss, 2008). All told, these findings suggest the fruitfulness of conceptualizing congruence in accordance with the correspondence principle.

Conclusions

The motive profile approach to the PSE we have presented here is not entirely new. As we have shown, looking at picture profiles was something that McClelland and colleagues (1953) already had on their minds when they started work on PSE motive measures, and it was further explicated by Atkinson (1958b) in his social-learning account of differential response patterns to pictures. Subsequently, though, researchers lost sight of this aspect of the PSE and got bogged down by intricate theorizing (e.g., Atkinson & Birch, 1970) and skirmishes with critics from the classical test theory camp (e.g., Entwisle, 1972; Lilienfeld et al., 2000; McClelland, 1980). Taking their cue from modern personality theory (Mischel & Shoda, 1995), Schultheiss et al. (2008) and Busch and Hofer (2012) have only very recently redirected attention from PSE motive sum scores to individual picture scores in an attempt to understand why PSE motive scores can have substantial retest reliability in the absence of internal consistency. As we have shown, conceptualizing the PSE in terms of particular pictures and responses has numerous additional implications for understanding how implicit motives work.

Perhaps some readers are concerned that our “molecular” approach to the PSE sacrifices the unity of implicit motives while opening a Pandora's box of post-hoc theorizing. We disagree with this argument for three reasons. First, remember that PSE motive measures are not the motives themselves. They are only the lens through which we can recognize and measure a person's fundamental motivational needs. This lens is multi-faceted, necessarily shaped by individual learning histories in terms of the cues that are associated with need satisfaction and the responses that are instrumental for it. Just as the fundamental need to eat can be expressed through diverse culinary customs and satisfied with all kinds of food, so can the unitary needs for achievement, power, or affiliation express themselves in a variety of behaviors to a variety of cues. The PSE simply reflects this fact and can be decomposed accordingly.

Second, if motive researchers use experimental motive arousal to study changes in imaginative stories, they can dissect which types of arousal lead to which types of responses to which type of cues (see McClelland et al., 1953). They can thus identify the causal roots of why some individuals respond more to one type of picture than the other or why certain motivational responses come up more frequently under some circumstances than others.

Applied to comparisons between actual versus imagined behavior in response to situational versus picture cues, this approach would also allow one to test the correspondence hypothesis.

Third, we simply think it is wrong for us as scientists to close our eyes to the peculiarities and problems of our measurement instruments. It is our impression that the field of implicit motive research has stagnated unduly because of unfruitful, abstract debates about motive measurement in the absence of substantial research on the instrument itself. By opening this black box and exploring its inner workings, we might actually learn a thing or two about how the PSE works and how motives shape fantasies, feelings, and behavior.

Acknowledgements

We thank our colleagues Martin Köllner and Andreas G. Rösch for helpful comments on a previous version of the paper. Portions of the data we report here were collected with support by a Deutsche Forschungsgemeinschaft (DFG) grant (SCHU 1210/3-1).

Short Biographies

Oliver C. Schultheiss received his diploma (1994) and PhD (1996) from Friedrich-Alexander University, Erlangen, Germany. From 1997 to 1999, he had his post-doc at Harvard University, Boston, USA and University of Potsdam, Potsdam, Germany. He was an assistant professor (2000–2005) and associate professor of psychology (2005–2007) at the University of Michigan, Ann Arbor, USA. Since 2007, he is a professor of psychology at Friedrich-Alexander University, Erlangen, Germany. His current areas of research include endocrine correlates of implicit motives; effects of implicit motives on brain activation, cognitive, and behavioral responses to perceived facial expressions of emotion; and factors influencing and outcomes influenced by the interplay of implicit motives and explicit goals and values. For more information, please visit the Human Motivation and Affective Neuroscience Laboratory website at <http://www.psych2.phil.uni-erlangen.de/~oschult/humanlab/index.htm>

Maria Schultheiss received her bachelor's (BSc, 2010) and master's (MSc, 2012) degrees from Friedrich-Alexander University, Erlangen, Germany and is currently enrolled in a clinical psychology/cognitive-behavioral therapy program. Her current areas of research include utilization of implicit and explicit motive measures in clinical diagnosis and the therapeutic process, developing interventions for enhancing congruence between implicit and explicit motivational needs, and behavioral assessment of parent-child attachment.

Note

* Correspondence: Department of Psychology, Nägelsbachstrasse 49b, Friedrich-Alexander University, 91052 Erlangen, Germany. Email: oliver.schultheiss@fau.de

References

- Atkinson, J. W. (1958a). *Motives in Fantasy, Action, and Society: A Method of Assessment and Study*. Princeton, NJ: Van Nostrand.
- Atkinson, J. W. (1958b). Thematic apperceptive measurement of motives within the context of a theory of motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society: A method of assessment and study* (pp. 596–616). Princeton, NJ: Van Nostrand.
- Atkinson, J. W. (1981). Studying personality in the context of an advanced motivational psychology. *American Psychologist*, **36**, 117–128.
- Atkinson, J. W., & Birch, D. (1970). *The Dynamics of Action*. New York: Wiley.

- Atkinson, J. W., Bongort, K., & Price, L. H. (1977). Explorations using computer simulation to comprehend thematic apperceptive measurement of motivation. *Motivation and Emotion*, **1**, 1–27.
- Atkinson, J. W., Heyns, R. W., & Veroff, J. (1958). The effect of experimental arousal of the affiliation motive on thematic apperception. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society: A method of assessment and study* (pp. 95–104). Princeton, NJ: Van Nostrand.
- Auld, F., Jr. (1954). Contributions of behavior theory to projective testing. *Journal of Projective Techniques*, **18**, 421–426.
- Bakan, D. (1966). *The duality of human existence: Isolation and communion in Western man*. Boston MA: Beacon Press.
- Blankenship, V., & Zoota, A. L. (1998). Comparing power imagery in TATs written by hand or on the computer. *Behavior Research Methods, Instruments, & Computers*, **30**, 441–448.
- Borsboom, D. (2005). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge, UK: Cambridge University Press.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, **111**, 1061–1071.
- Brunstein, J. C. (2010). Implicit motives and explicit goals: The role of motivational congruence in emotional well-being. In O. C. Schultheiss, & J. C. Brunstein (Eds.), *Implicit motives* (pp. 347–374). New York: Oxford University Press.
- Brunstein, J. C., Schultheiss, O. C., & Grässmann, R. (1998). Personal goals and emotional well-being: The moderating role of motive dispositions. *Journal of Personality and Social Psychology*, **75**, 494–508.
- Busch, H., & Hofer, J. (2012). Picture story exercise picture set in a German and a Cameroonian sample: Ipsative stability, retest reliability, and sample-level stability. *European Journal of Psychological Assessment*, **28**, 125–131.
- Entwisle, D. R. (1972). To dispel fantasies about fantasy-based measures of achievement motivation. *Psychological Bulletin*, **77**, 377–391.
- Fleming, J. (1982). Projective and psychometric approaches to measurement: The case of fear of success. In A. J. Stewart (Ed.), *Motivation and society. A volume in honor of David C. McClelland* (pp. 63–96). San Francisco, CA: Jossey-Bass.
- Fleming, J., & Horner, M. S. (1992). The motive to avoid success. In C. P. Smith (Ed.), *Motivation and Personality: Handbook of Thematic Content Analysis* (pp. 179–189). New York: Cambridge University Press.
- Furr, R. M. (2009). Profile analysis in person–situation integration. *Journal of Research in Personality*, **43**, 196–207.
- Goss, A. E., & Brownell, M. H. (1957). Stimulus–response concepts and principles applied to projective test behavior. *Journal of Personality*, **25**, 505–523.
- Heckhausen, H. (1963). *Hoffnung und Furcht in der Leistungsmotivation [Hope and Fear Components of Achievement Motivation]*. Meisenheim am Glan: Anton Hain.
- Hofer, J., & Busch, H. (2013). Living in accordance with one's implicit motives: Cross-cultural evidence for beneficial effects of motive–goal congruence and motive satisfaction. In A. Efklides, & D. Moraitou (Eds.), *A Positive Psychology Perspective on Quality of Life* (pp. 51–66). Dordrecht, Netherlands: Springer.
- Jackson, D. N. (1984). *Personality Research Form* (3rd edn). Port Huron, MI: Sigma Assessment Systems, Inc.
- Kazén, M., & Kuhl, J. (2011). Directional discrepancy between implicit and explicit power motives is related to well-being among managers. *Motivation and Emotion*, **35**, 317–327.
- Lilienfeld, S. O., Wood, J. E., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest*, **1**, 27–66.
- Lundy, A. (1985). The reliability of the Thematic Apperception Test. *Journal of Personality Assessment*, **49**, 141–145.
- McAdams, D. P. (1980). A thematic coding system for the intimacy motive. *Journal of Research in Personality*, **14**, 413–432.
- McAdams, D. P., & Powers, J. (1981). Themes of intimacy in behavior and thought. *Journal of Personality and Social Psychology*, **40**, 573–587.
- McClelland, D. C. (1958). Methods of measuring human motivation. In J. W. Atkinson (Ed.), *Motives in Fantasy, Action, and Society: A Method of Assessment and Study* (pp. 7–42). Princeton, NJ: Van Nostrand.
- McClelland, D. C. (1980). Motive dispositions: The merits of operant and respondent measures. In L. Wheeler (Ed.), *Review of Personality and Social Psychology* (Vol. **1**, pp. 10–41). Beverly Hills, CA: Sage.
- McClelland, D. C. (1987). *Human Motivation*. New York: Cambridge University Press.
- McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). *The Achievement Motive*. New York: Appleton–Century–Crofts.
- McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review*, **96**, 690–702.
- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, **102**, 246–268.
- Morgan, C., & Murray, H. A. (1935). A method for investigating fantasies: The Thematic Apperception Test. *Archives of Neurology and Psychiatry*, **34**, 289–306.
- Pang, J. S., & Schultheiss, O. C. (2005). Assessing implicit motives in U.S. College students: Effects of picture type and position, gender and ethnicity, and cross-cultural comparisons. *Journal of Personality Assessment*, **85**, 280–294.

- Ramsay, J. E., & Pang, J. S. (forthcoming). Set ambiguity: A key determinant of reliability and validity in the picture story exercise. *Motivation and Emotion*.
- Reuman, D. A. (1982). Ipsative behavioral variability and the quality of thematic apperceptive measurement of the achievement motive. *Journal of Personality and Social Psychology*, **43**, 1098–1110.
- Schultheiss, O. C. (2008). Implicit motives. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of Personality: Theory and Research* (3rd edn. pp. 603–633). New York: Guilford.
- Schultheiss, O. C. (2013). The hormonal correlates of implicit motives. *Social and Personality Psychology Compass*, **7**, 52–65.
- Schultheiss, O. C., & Brunstein, J. C. (2001). Assessing implicit motives with a research version of the TAT: Picture profiles, gender differences, and relations to other personality measures. *Journal of Personality Assessment*, **77**, 71–86.
- Schultheiss, O. C., & Brunstein, J. C. (2002). Inhibited power motivation and persuasive communication: A lens model analysis. *Journal of Personality*, **70**, 553–582.
- Schultheiss, O. C., & Brunstein, J. C. (Eds.), (2010). *Implicit Motives*. New York: Oxford University Press.
- Schultheiss, O. C., & Pang, J. S. (2007). Measuring implicit motives. In R. W. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 322–344). New York: Guilford.
- Schultheiss, O. C., Campbell, K. L., & McClelland, D. C. (1999). Implicit power motivation moderates men's testosterone responses to imagined and real dominance success. *Hormones and Behavior*, **36**, 234–241.
- Schultheiss, O. C., Liening, S., & Schad, D. (2008). The reliability of a Picture Story Exercise measure of implicit motives: Estimates of internal consistency, retest reliability, and ipsative stability. *Journal of Research in Personality*, **42**, 1560–1571.
- Schultheiss, O. C., Patalakh, M., Rawolle, M., Liening, S., & MacInnes, J. J. (2011). Referential competence is associated with motivational congruence. *Journal of Research in Personality*, **45**, 59–70.
- Schultheiss, O. C., Wirth, M. M., Torges, C. M., Pang, J. S., Villacorta, M. A., & Welsh, K. M. (2005). Effects of implicit power motivation on men's and women's implicit learning and testosterone changes after social victory or defeat. *Journal of Personality and Social Psychology*, **88**, 174–188.
- Schultheiss, O. C., Yankova, D., Dirlikov, B., & Schad, D. J. (2009). Are implicit and explicit motive measures statistically independent? A fair and balanced test using the picture story exercise and a cue- and response-matched questionnaire measure. *Journal of Personality Assessment*, **91**, 72–81.
- Shoda, Y., Mischel, W., & Wright, J. C. (1994). Intraindividual stability in the organization and patterning of behavior: Incorporating psychological situations into the idiographic analysis of personality. *Journal of Personality and Social Psychology*, **67**, 674–687.
- Smith, C. P. (Ed.). (1992). *Motivation and Personality: Handbook of Thematic Content Analysis*. New York: Cambridge University Press.
- Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin*, **112**, 140–154.
- Stanton, S. J., & Schultheiss, O. C. (2007). Basal and dynamic relationships between implicit power motivation and estradiol in women. *Hormones and Behavior*, **52**, 571–580.
- Thrash, T. M., Maruskin, L. A., & Martin, C. C. (2012). Implicit-explicit motive congruence. In R. M. Ryan (Ed.), *The Oxford Handbook of Human Motivation* (pp. 141–156). New York: Oxford University Press.
- Tuerlinckx, F., De Boeck, P., & Lens, W. (2002). Measuring needs with the Thematic Apperception Test: A psychometric study. *Journal of Personality and Social Psychology*, **82**, 448–461.
- Veroff, J. (1957). Development and validation of a projective measure of power motivation. *Journal of Abnormal and Social Psychology*, **54**, 1–8.
- Winter, D. G. (1973). *The Power Motive*. New York: Free Press.
- Winter, D. G. (1991). Measuring personality at a distance: Development of an integrated system for scoring motives in running text. In D. J. Ozer, J. M. Healy, & A. J. Stewart (Eds.), *Perspectives in Personality* (Vol. 3, pp. 59–89). London: Jessica Kingsley.
- Winter, D. G. (1994). *Manual for Scoring Motive Imagery in Running Text* (4th edn). Ann Arbor: Department of Psychology, University of Michigan. Unpublished manuscript.
- Winter, D. G., & Stewart, A. J. (1977). Power motive reliability as a function of retest instructions. *Journal of Consulting and Clinical Psychology*, **45**, 436–440.
- Wirth, M. M., Welsh, K. M., & Schultheiss, O. C. (2006). Salivary cortisol changes in humans after winning or losing a dominance contest depend on implicit power motivation. *Hormones and Behavior*, **49**, 346–352.