Chapter 1

Introduction

PURPOSE AND PRESENTING PROBLEM

Human relations education is a mandatory component of the initial training curriculum for all new Air Force personnel. Enlisted members receive several hours of race and gender relations education as part of their first six weeks of Air Force Basic Military Training (AFBMT). Varying amounts of similar training hours are included in the Air Force's principal officer pre-commissioning programs (USAF Academy, AFROTC, Air Force OTS, etc.). Shorter forms of such training sessions are then repeated throughout the military careers of both officers and enlisted personnel — primarily in the first month of each geographical re-location or re-assignment.

In conducting this training, it is accepted that racism and sexism likely afflict new Air Force members to varying degrees — and that the degree of such affliction is largely the result of cultural/ environmental and developmental processes extending across the prior decades of each member's life. The instructors responsible for Air Force human relations training are not likely to expect radical changes in individual levels of sexual or racial prejudice after a few hours of training. As such, this training usually emphasizes the non-negotiable requirement to strictly comply with Air Force human relations standards — though this training will also include appeals for cultural sensitivity and an appreciation of diversity. Air Force human relations training likewise emphasizes the importance of fairness and non-prejudicial relations for successful team-building and unit cohesion — and the likely contribution of such to mission accomplishment. Still, the bottom line is the requirement for compliance with Air Force standards — and the dire consequences (both personal and organizational) for failing to comply.

Since the trainee populations are quite large (tens of thousands per year), potential validation of competing approaches for this training should be quite feasible if we limit the validation effort to readily measurable outcomes. Over a period of several years, the number of gender or racially-motivated infractions committed by first-term (i.e., first enlistment) airmen can be compared to similar incident rates in similar civilian populations. The Air Force can also look at changes expressed in its standardized organizational climate assessment surveys. However, such data are more likely to reflect *overt* and presumably *conscious* and *temporary* responses to (among other things) our commander's and supervisor's emphasis or dedication to human relations standards — as much as the changing content of our human relations training. While such studies might yield valuable insights, there are at least two significant deficiencies associated with such overt group-level approaches.

First of all, while the rates of gender and racially-flavored incidents might be reasonably low, the degree to which covert or non-conscious sexism and racism are a problem might be obscured. For instance, a race-based prejudice might operate habitually or non-consciously in a supervisor — conceivably affecting this supervisor's appraisal of selected subordinates.

If such a condition is common in the Air Force, members predicted to perform at equivalent levels might appear to be performing at different levels when annual performance appraisal data are sorted by race. Although there are equally-likely alternative explanations for such outcomes, this is exactly the sort of outcome found in a recent study of prediction lines (across races) when the Air Force Officer Qualifying Test (AFOQT) was used to predict future performance (Roberts, 1994). Specifically, black officers typically received lower performance ratings than non-black officers with equivalent pre-commissioning scores on the AFOQT.

If a covert or non-conscious racial prejudice *is* at work here, an incident-centered assessment of the impact of Air Force race relations education is unlikely to reflect it — and it follows that such an approach is not likely to reveal the impact of curriculum modifications on such mechanisms. An alternative methodology is suggested — one that may illuminate the effects of prejudicial expectations on private assessments of subordinate performance.

The second significant deficiency associated with a traditional overt group-level approach is related to the difficulty in *attributing* variation in sexism/racism-related incident rates to varying training interventions (mentioned above). Obviously, these incident rates can change in response to any number of things — such rates are likely to reflect the changing attitudes expressed by the Air Force leadership. Increasing incident rates at the squadron or wing level are likely to evoke a response from local commanders — after which, incident rates are likely to drop. Pentagon or senior leadership emphasis might change in response to incident rates, organizational climate assessments, media coverage or changes in the political climate. Therefore, changes in race relations education content (or even the presence or absence of such education) may explain relatively little variance in race-related incident rates. Common sense dictates that the elimination of race relations education in the military is not an option. However, if we are to look for an empirically-based validation of possible improvements to such training, we will need strategies that tie apparent changes in the trainees directly to the training — minimizing confounds due to changing societal or organizational priorities.

The purpose of the multi-phase research program described here-in is the development of alternative validation techniques (for assessing the effectiveness of competing approaches to Air Force human relations education) — emphasizing the measure of internal, habitual and/or covert processes within individual trainees — especially processes thought to bias judgments of others' performance. This research program is divided into phases — where the specifics of the later phases depend upon the results of their predecessors. The first (pre-laboratory phase) centers on the development and selection of survey items that can reliably measure levels of *prejudicial expectation* (PE) largely apart from levels of authoritarianism (sometimes called fascism), social desirability, racial integration/equal opportunity attitudes and cultural/developmental *reinforcement for racist judgments* (RRJC). The latter construct (RRJC) derived from Karl Pribram's objections (personal communication, 1994) concerning the effects of upbringing on cross-racial judgments.

The second phase (our laboratory phase) embeds these few items in a large (and largely) deceptive survey to predict racial prejudice effects on information processing of minority-target performance data — and then tests these predictions in a laboratory study. Instead of the traditional evaluative traits often used in person-memory research (friendly/unfriendly or likable/non-likable), this laboratory effort uses a hypothetical annual appraisal dimension we call *academic potential* — in the hope that our academically-invested subjects will feel qualified to assess such a trait dimension. In other words, instead of assessing a target's affability, our subjects will assess the target's academic prowess — much as they might size up the competition in each class during the first week of a semester. The intent is to see how racial prejudices will compare to the previously found effects of experimenter-provided expectancies — and to see if the recall effects of evaluative person-concepts (e.g., friendly/likable) are replicated when using a less personal or less social trait like academic potential. The goal here is to move from affability to a more job performance-related dimension while keeping our raters in a realm in which they can make somewhat expert judgments.

The third (post-dissertation phase) involves the further development and confirmatory analysis of survey items that can measure somewhat distinct facets of racism and some of its likely developmental precursors — a few of which may be more responsive to educational interventions than others. The fourth phase (a moderate post-doctorate effort) will use data collected in the first three phases to assess the possible relationships between facets of racism and the person-memory to person-judgment link. Finally, the fifth phase (a major post-doctorate effort) will involve substantial Air Force man-hours in a hybrid field and laboratory study (hopefully) validating the race relations component of AFBMT.

The last three phases (all of which are viewed as post-dissertation activities) will be proposed to the Air Force Office of Scientific Research and the Research Division of the Defense Equal Opportunity Management Institute. If approved for funding, these phases will be executed in partnership with Air Force social psychologists specializing in attitude change interventions and Air Force cognitive psychologists experienced with neurophysiological manifestations of attitude and emotion. Although these later phases expand the scope of our efforts (to include, for example, recall-judgment relationships, structural equation modeling of developmental facets of racism, as well as electro-physiological responses to racially-noxious imagery), the central interest of the entire research program remains with racism's possible effects on person-memory processes. The dissertation phases of this effort deal exclusively with such effects. The processes of attitude change are not dealt with at all in the dissertation portion of my efforts — even tangentially — although person-memory processes are likely to play a major role in racial attitude development and maintenance. Substantial literature reviews will have to be performed (especially related to attitude development and change-seeking interventions). That said, we hope this research program will improve the Air Force's ability to compare different approaches to race-related attitude change.

CONCEPT DEFINITIONS: STEREOTYPES VERSUS PREJUDICE

Racial stereotypes have traditionally been viewed as social categorizations created to reduce the complexity of social perceptions (Allport, 1954 and Hamilton, 1979, cited in Devine and Baker, 1991, and Taylor, 1978). However, this tradition replaced a much older one whereby racial stereotyping was viewed as an irrational or faulty "reasoning process rigidly unresponsive to feedback" (Taylor, 1988, p. 778, citing Lippmann, 1922).

Unfortunately, by-products of Allport's social categorization process can intensify racial prejudice. Taylor, et al (1978), conducted three studies where subjects observed audio-visual portrayals of small multi-racial groups. Testing person-memory recall behaviors, she found that subjects encode person information by racial groups and gender and that this process causes subjects to minimize within-group differences and exaggerate between-group differences. In other words, once categorized in a group, an individual is assumed to share characteristics with other group members. This, in turn, can affect how others perceive, evaluate, and react or behave towards the individual (Devine and Baker, 1991).

But Devine (1989) argues that there is a key distinction between racial prejudice and acting on racial stereotypes. She points to evidence that racial stereotypes are well established in children's memories before the children are mature enough to question their validity. Later in life, the adolescent or

4

grown-up may challenge the stereotype with new personal beliefs, but the stereotype has a longer history of practiced activation, and is therefore more accessible than the newer personal beliefs. Devine believes this explains why both high and low racists (reflecting subjects with different personal beliefs) are equally likely to summon up stereotypical character assessments when judgment tasks do not allow time for conscious suppression of these simplistic categorizations.

Bodenhausen (1988) found that our use of stereotype categorization in person memory does not really influence the *impression* incoming data makes on our judgments, but instead, it biases our processing of judgment-relevant data via selective emphasis of stereotype-consistent data over inconsistent data. This suggests better recall of stereotype-consistent information. This also supports Taylor and Crocker's model whereby schematic processing filters out stereotype-inconsistent data along with irrelevant data (1981, cited in O'Sullivan, et al, 1984).

O'Sullivan (1984) points out that increased recall of inconsistent information is supported by numerous studies (e.g., Hastie and Kumar, 1979, Srull, 1981) but that other studies have found that stereotype consistent data are remembered better (e.g., Bellezza and Bower, 1981, Cohen, 1977). On the other hand, Wyer and Gordon (1982, cited in O'Sullivan, et al, 1984) assert that inconsistent information is distinctive, and therefore, deeply processed and well-recalled. This suggests that inconsistent data should be effective in weakening stereotype-supported prejudice (O'Sullivan, et al, 1984). However, O'Sullivan and Durso's (1984) results "showed that the introduction of information that is highly incongruent with a schema makes schematic information more memorable" (p. 55) just as Srull (1981) had found that adding a higher proportion of inconsistent data actually increased recall of stereotype consistent data. They suggest that this may "be one mechanism by which inappropriate schemata persevere in the face of counter-evidence" (p.55) — in other words, they're suggesting that stereotype-disconfirming evidence can sometimes act as though it is stereotype enhancing! They found their results to be roughly consistent with the Hastie associative network model. This interpretation is also consistent with Hamilton's (1989) study of the H-SAN model.

Stangor and Ruble (1989) may have cleared up some of this confusion. Their study demonstrated that the type of information we remember best (consistent or inconsistent) depends partially on how much we already know about the target person when the new information comes our way. If we know very little, then "this may lead to elaborative-processing of expectancy-disconfirming information in an attempt to......form a coherent representation" of the target (p. 19). Under such circumstances, inconsistent information forms more inter-item linkages in memory "because it is more thoroughly processed" (p. 19). But once expectancies are well-established, they may act as retrieval cues for consistent data, and as filters for inconsistent data — presenting such data as situational artifacts, or they may simply distort inconsistent data so that it appears consistent. Partially because of the mixed results just cited, our designs also look at other evidence of stereotypical cognitive structure (e.g., inter-recall intervals) beyond just preferential recall of consistent versus inconsistent information.

If Devine's interpretation of the research is correct, cognitive schemes built on racial stereotypes are likely to be quite resilient and indifferent when challenged by traditional race relations training, while personal beliefs in this area may be at least moderately malleable. The results from our five phases may ultimately allow us to assess the degree to which Air Force race relations training differentially alters the effects of racial stereotypes and racial attitudes.

CONCEPT DEFINITIONS: RACE RELATIONS TRAINING

Criteria for successful race relations training suggested by Foeman (1991) and Katz and Ivey (1978) engender a mildly skeptical view of existing Air Force training. Foeman lists several elements usually included in legitimate race relations training programs. These include: demystification of racism and race-related social dynamics, articulation of other racial group perspectives, guided self-examination of these perspectives, identifying the valid components of these minority perspectives, and using these new insights to solve problems, adjust expectations and change behaviors. Foeman (1991) describes the following goals for race relations training:

1) Establishing a relevant information base for the trainees;

2) Encouraging trainee-ownership of racial attitudes and behaviors;

3) Increasing cross-racial dialogue and experience, hopefully replacing less integrated and socially-constructive patterns of behavior.

Air Force race relations training certainly attempts to meet goals 1 and 2, but goal 3 may largely be an outcome of training, living and working within the Air Force's multi-racial workforce. However, since movement towards goal 3 is unstructured, goal attainment may be less rapid and/or less reliable.

U.S. Air Force initial race relations education contains nine hours of formal instruction and discussion. These nine hours are integrated into much larger curricula and are often split between USAF Basic Training and the Social Actions First Duty Station orientations. The former is an intensive six-week introduction to military leadership environments and Air Force requirements. The latter is a locally-provided, yet standardized, training session encountered by new Airmen upon arrival at their first post-training assignment. For purposes of our research program, we intend to have the entire nine hours included in the first six-weeks of training. However, either way it includes the same components — three hours of lecture on the historical, cultural and social bases of racism, three hours of disturbing and thought-provoking films and guided discussion, and three hours of lectures on Air Force standards, race discrimination complaint procedures, racial incident investigation procedures, and relevant court-martial procedures and typical consequences or sanctions from such legal procedures.

Beyond the success measured via the Air Force's own comprehension/retention or *knowledge* tests, the literature provides some reasons to expect at least moderate degrees of effectiveness for USAF race relations training. Sedlacek, et al (1978) found significant improvements in racial attitudes after as little as two hours of race relations training. They competed different training approaches (role-plays, disturbing films, and a guided discussion following a path similar to Foeman's recommended elements listed above). Unfortunately, their measures were very likely subject to social desirability response sets. However, given that the Air Force's training can include much of what is contained in two of Sedlacek's three approaches, combining this material with a terse lecture on USAF standards and consequences should be expected to have some effect. It appears to be reasonable for the Air Force to expect some measure of benefit, despite the largely didactic approach of their race relations training.

However, Foeman (1991) criticizes this didactic approach to race-relations training. While conceding that deficiencies in cross-race interpersonal relations (CRIRs) may be related to knowledge (or lack of knowledge) problems, she cites evidence that achieving abilities needed for healthy CRIRs requires real-life role-models (viewed in action) and challenging exercises for the students. Concise lecture-formats of the didactic approach, she argues, are unlikely to provide such experiences. Kochman (1982, cited in Foeman, 1991) suggests that this approach appeals the most to middle-class whites, and is likely to result in a well-informed cavalier who can "be sensitive to *others* as they are brought into the *mainstream*.......eliciting a somewhat patronizing posture from historically privileged listeners rather than a functional dialogue among participants" (p. 258, original author's emphasis conveyed). Foeman also cites previous evidence that didactic race-relations courses can make racial attitudes *worse!* (Tansik and Driskill, 1977).

As an alternative, Foeman says her survey of the field found far greater support for an experiential approach. The studies of Bernard and Ben (1988, cited in Foeman, 1991) found that cross-racial acceptance was much less dependent on value or belief congruence than on personal contact with members of another race. They also found that attitude shifts derived from this personal multi-racial contact were more likely to persist six-weeks after the multi-racial training experience was completed. A study by Pate (1981, cited in Foeman, 1991) also supports this assertion. A survey by Katz and Ivey (1977) found some successes in training programs using inter-racial encounters, but no successes in purely knowledge-based programs. Williams and Giles (1992) found mixed results, even for experiential prejudice-reduction training.

CONCEPT APPLICATION

M. R. Dansby, Director of Research at the Defense Equal Opportunity Management Institute (DEOMI) reports that formal evaluation of the effectiveness of Air Force race relations training (traditionally provided for new recruits) has not yet been performed. In general, research on the effectiveness of race relations training is apparently very limited (Sedlacek, et al, 1978). This may be "due in part to the emotionally laden nature of the issue, and in part to the difficulties inherent in the selection of appropriate criteria to assess program effectiveness" (p. 196). It is probably *related* to the frequent use of measurement instruments with overwhelming social desirability response problems (e.g. Katz and Ivey, 1978, and Sedlacek, et al, 1978) — and it may also be related to the frequent use of student *post-training satisfaction* questionnaires as the only measure of training effectiveness (Williams and Giles, 1992).

I would like to argue that the "emotionally laden nature" of measuring one's racism is part of the reason that successfully deceptive/clandestine racism measures are hard to come by — our subjects' sensitivities, and therefore, their suspicions, are likely to be elevated. It also makes the use of tests plagued by social desirability especially problematic when using racism as an independent variable in post-test settings.

For these reasons, I am proposing the use of racism assessments based on person-memory recall behavior measures in the context of the Hastie-Srull associative network model (described below). Previous arguments and at least one experimental study have suggested that both high and low

prejudice persons are equally knowledgeable of racial/cultural stereotypes (Devine, 1989). This study found that both high and low prejudice persons produced stereotype-congruent evaluations of ambiguous behaviors if their capacity to consciously monitor stereotype activation was constrained. Furthermore, this study demonstrated that only low prejudice subjects consistently inhibited stereotype-congruent thoughts and replaced them with stereotype-contrary thoughts reflecting equality. These findings imply that all human beings are afflicted with some capacity (or tendency) to commit racially-determined pre-judgments, with low prejudice subjects *inhibiting* stereotype-congruent thoughts.

If the above conclusion is over-generalized, evidence for this error might surface while testing the Hastie-Srull associative network (H-SAN) model with high and low prejudice subjects. Previous studies have assessed the importance of congruence in recall tasks related to person-memory (Hastie, 1979). In these studies, subjects read a list of eight synonymous adjectives describing a single personality trait for a fictional character. They then read a series of sentences describing actions or events involving this fictional character — some of which were congruent and some incongruent with the single personality trait (the initial impression of the character formed by reading the original adjective list).

The findings from these studies (and *some* more recent studies cited in Hamilton, et al, 1989) provide support for the H-SAN model as depicted in Figure 1. In these studies, subjects typically are asked to recall what they can from the lists of event-describing sentences. Three principal measures result from such data: probability of recall, interrecall interval durations, and sequence of recollections.

For studies involving a single personality trait, all three measures support the H-SAN model. Incongruent events are more likely to be recalled. Consistent with the lack of direct interitem connections between encodings of congruent behaviors (again see Figure 1), interrecall interval durations between successively recalled congruent behaviors are longer. Given that a congruent item has just been recalled, the next item recalled is likely incongruent; however, after recalling an incongruent item, the following is equally likely to be of either type.
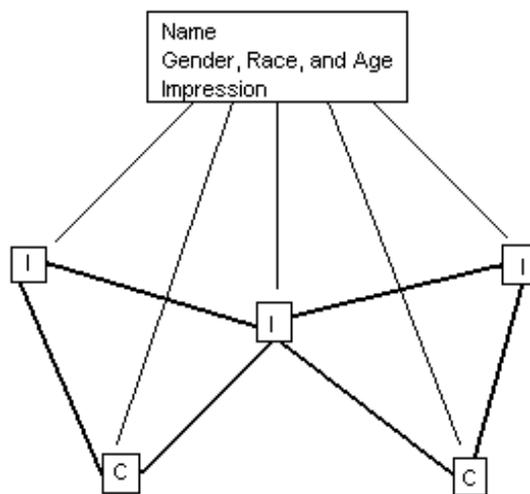

Figure 1. Hastie-Srull Associative Network (H-SAN) with Congruencies and Incongruencies

8

The laboratory portion of Phase II of this research program evaluates the degree to which racial prejudice is manifested by pre-existing expectancies (for memories of racial minority targets). This effort includes assessment of conditional probabilities in recall sequences, interrecall interval durations, and quantity and nature of items recalled. The study attempts to demonstrate that subjects afflicted with a greater degree of racial prejudice show signs of stereotype-centered cognitive structure before other subjects. Such outcomes (if fully obtained) would support the notion that low prejudice subjects lack the same degree of relevant pre-existing structure as high prejudice subjects — a finding contrary to at least the broadest interpretation of the Devine studies.

The use of associative-network studies may provide valuable insights in unobtrusively assessing racial prejudice. Unless subjects are thoroughly familiar with the work of Hastie and Srull (including their experimental designs and data analysis protocols), they should find it very difficult to corrupt their own data — even if they know their racism tendencies are being assessed. They might, for example, consciously increase the number of stereotype-inconsistent scenes or descriptors that they recall right up front, so as to appear open-minded or egalitarian. However, one interpretation of the H-SAN model would predict exactly this sort of recall behavior for high-racism scorers (as compared with normal recency and primacy effects).

Use of the H-SAN model in the hybrid field and lab experiment (of the fifth phase) is intended to build on the results of the work done in preceding phases. Because this specific sort of application for the H-SAN model is yet to be validated, the five-phased research program alluded to here-in includes the use of other moderately-veiled racism measures (both paper/pencil and physiological). This program's ultimate purpose is to assess (via multiple techniques) the *effectiveness* of standard Air Force race relations training, and at the same time, to assess the validity of using H-SAN techniques for assessments of one possible manifestation of such *effectiveness*.