



### Additional Panelists



 Art Gutman, Professor of Psychology, I-O graduate program at Florida Institute of Technology.



- David Copus, distinguished employment attorney with Morristown, New Jersey office of Ogletree Deakins.
- James Outtz, consulting expert or testifying expert for plaintiffs and defendants in litigation.
- James C. Sharf, Sharf & Associates, Employment Risk Advisors, Inc.



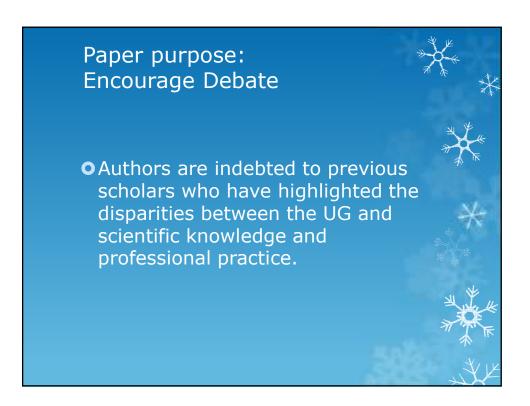
### McDaniel's 10 Minutes





- McDaniel, M.A., Kepes, S., Banks, G. C. (in press). The *Uniform Guidelines* are a detriment to the field of personnel selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice.*
- Paper available on the SIOP web site.
- Encourage commentaries to be submitted



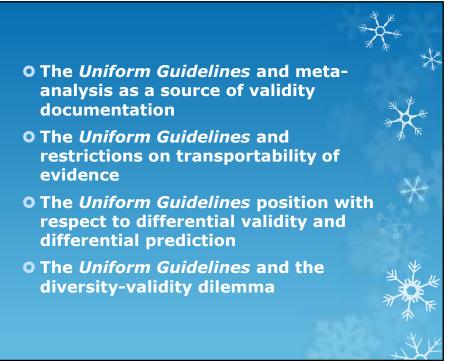








- O The *Uniform Guidelines* and evidence for validity based on content similarity
- O The *Uniform Guidelines* and evidence for validity based on construct validity
- O The *Uniform Guidelines* and its 1950's perspective on separate "types" of validity



O The *Uniform Guidelines* and false assumptions concerning adverse impact



- O We suggest that an implicit assumption of the Uniform Guidelines is that adverse impact is an indication of a flawed test.
- O We offer the alternative hypothesis that employment tests are an accurate assessment of subgroup differences in jobrelated attributes.
- O Adverse impact is the norm and not the exception.
- O Mean racial differences in employment test scores will be present for a very long time.

## Science and Federal Regulations



• Many Federal regulatory agencies do a good job of keeping regulations consistent with scientific findings.



- Many professional organizations are active in advising Federal regulators on changes needed in regulations to bring them into consistency with scientific knowledge.
- Until March of this year, that has not been the case with SIOP which had been silent on the issue for its 26 years of existence.



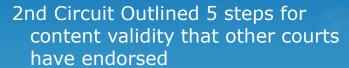






# McDaniel (and many others) Point to Following UGESP Excerpt • A selection procedure based on inferences about mental processes cannot be supported solely or primarily on the basis of content validity. Thus, a content strategy is not appropriate for demonstrating the validity of selection procedures which purport to measure traits or constructs such as intelligence, aptitude, personality, common sense, judgment, leadership and spatial ability. • No court has ever upheld this guidance ... in fact

## Guardians v. Civil Service (1980)



- (1) suitable job analysis
- (2) reasonable competence in test construction
- (3) test content related to job content
- (4) test content representative of job content
- (5) scoring systems selecting applicants who are likely to be better job performers.



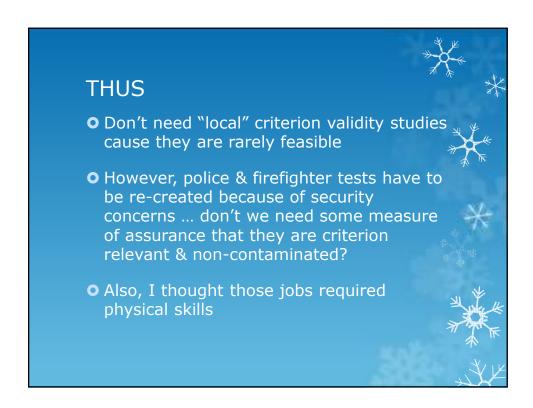
### Other Examples

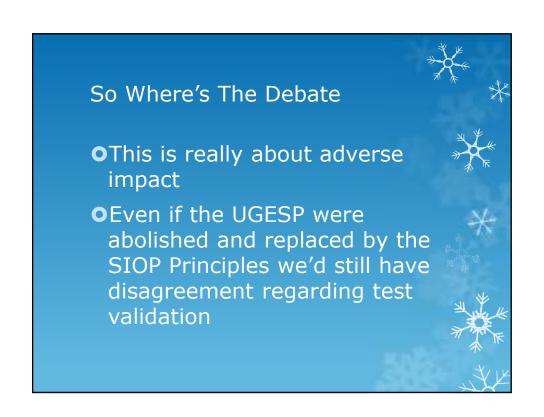
- In *Gillespie v Wisconsin* (1985), the 7<sup>th</sup> Circuit ruled neither the Uniform Guidelines nor the psychological literature express a blanket preference for criterion-related validity
- In *Police Officers v Columbus (1990)* the 6<sup>th</sup> Circuit, citing the 1987 SIOP Principles ruled that it is critical that selection instruments measure a substantial and important part of the job reliably, and provide adequate discrimination in the score ranges involved.
- Content validity supported in many subsequent cases, including the 2<sup>nd</sup> Circuit in *Gulino v. NY State*, 2006).



8



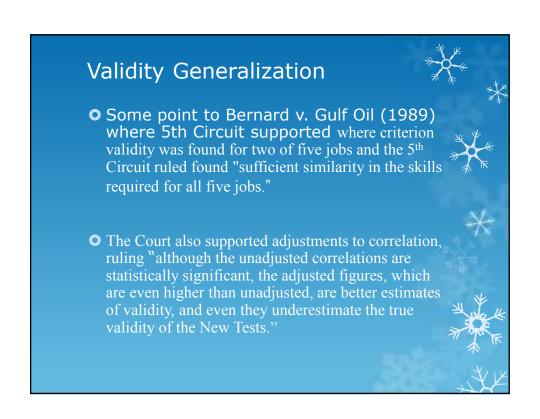




# Sharf Believes • Adverse impact was judicially created and never debated in CRA-91 • So what - do we declare all judicial law unconstitutional (e.g., rules for disparate treatment & mixed motive cases)







### On the Other Hand

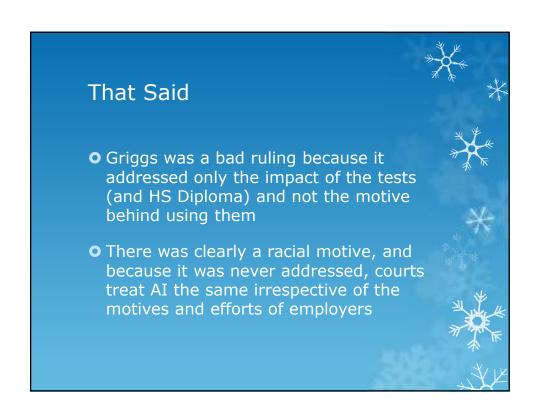
- In EEOC v. Atlas (1989), 6th Circuit rules "Atlas' validity generalization theory ignores the teachings of Albemarle by implying that no linkage or similarity between those jobs which had been previously researched by Hunter and those at the Company need be shown. The similarity was simply assumed."
- Similar ruling in Lewis v. Chicago (2005), but will leave for Ouutz who was an expert in that case.
- Also ... Mike Campion was expert on other side, and perhaps he can restage the debate right here, right now,

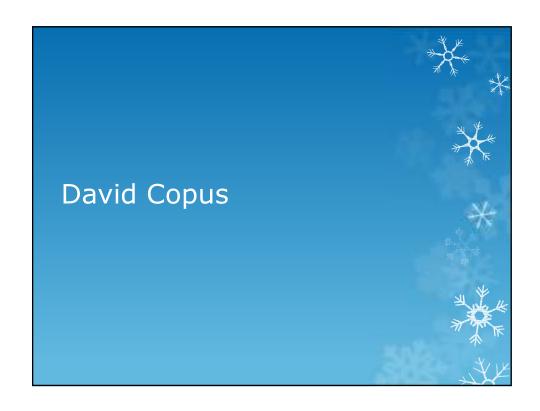
### **Multiple Choice Tests**

- As an academician, it amazes me how much weight is accorded to MC tests.

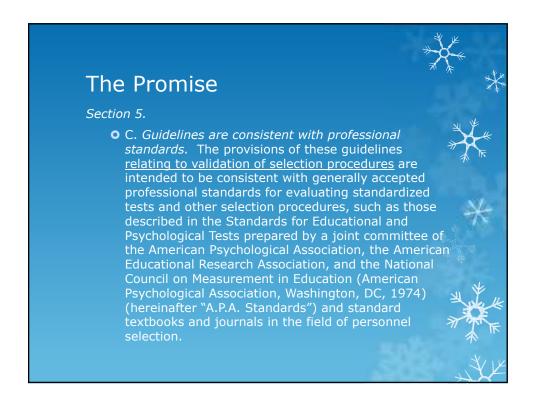
  Those of us who teach, especially graduate students (e.g., me & McD) would never place that much reliance on such tests to separate average from superior students
- I think there's something to be said for alternatives with less adverse impact ... but once again I will defer to Ouutz on that one because he has factual data on this issue

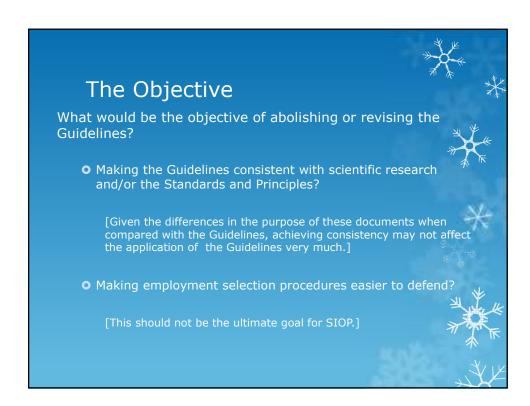


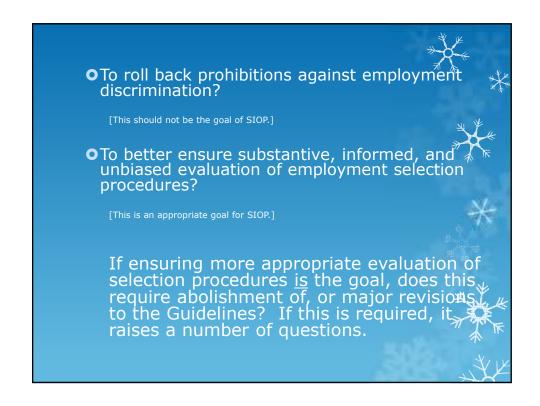


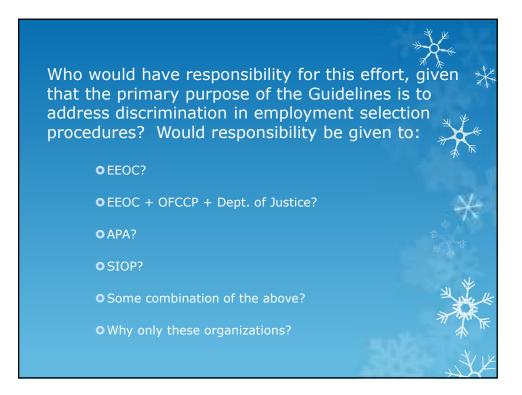




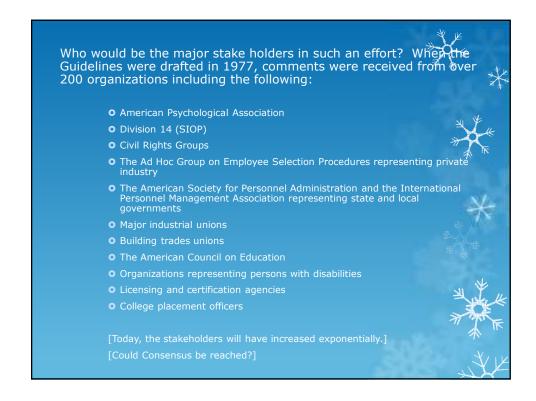


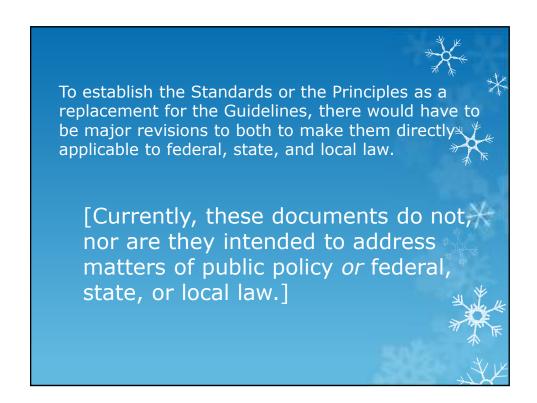












## Example of the Problem

A key provision of the Guidelines is that a search for alternatives must be part of any validation study where adverse impact exists.

• How would this provision be addressed if either the Standards or the Principles were to replace the Guidelines?

Would the Standards and Principles have to be merged into one document in preparation for replacing or serving as the basis for major revisions to the Guidelines?

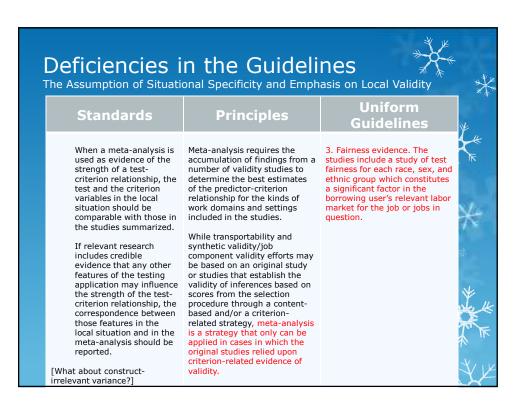
If these documents had to be merged:

- Who would be responsible for doing this?
- What would be the process?
- Who would be the stakeholders?

Here is a brief look at the level of consistency between the Standards, Principles, and Guidelines on key points targeted for discussion by this panel.

### Purpose of the Standards, Principles, and Uniform Guidelines **Standards Principles** The Standards makes no The Principles is intended to be The Guidelines are intended to attempt to provide psychometric consistent with the Standards. establish a uniform Federal answers to questions of public position in the area of prohibiting discrimination in employment practices on policy regarding the use of tests. Federal, state, and local In general, the Standards statutes, regulations, and case advocates that, within feasible grounds of race, color, religion, law regarding employment decisions exist. The *Principles* is sex, or national origin. limits, the relevant technical information be made available not intended to interpret these so that those involved in policy statues, regulations, and case The fundamental principle decisions may be fully informed. law, but can inform decision underlying the Guidelines is that making related to them. employer policies or practices The Standards do not attempt to which have an adverse impact repeat or incorporate the many legal This document is intended to be on employment opportunities of or regulatory requirements that might be relevant to the issues they address. In some areas, such as the collection, aspirational and to facilitate and any race, sex, or ethnic group assist the validation and use of are illegal under title VII and the Executive order unless justified selection procedures. analysis, and use of test data and results for different subgroups, the law may both require participants in the testing process to take certain actions and prohibit those participants from by business necessity. It is not intended to be mandatory, exhaustive, or definitive, and may not be taking other actions. Where it is apparent that one or more standards applicable to every situation. or comments address an issue on which established legal requirements may be particularly relevant, the standard, comment, or introductory material may make note of that fact. Lack of specific reference to legal requirements, however, does not imply that no relevant requirement exists.

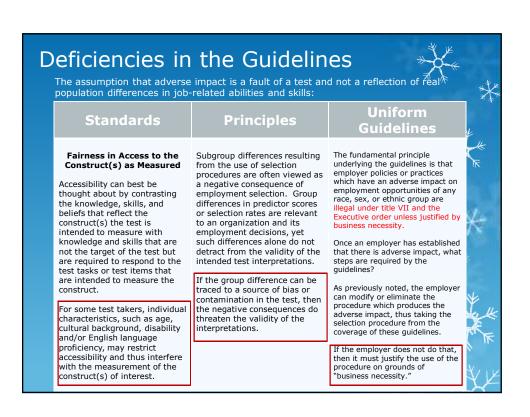
### Deficiencies in the Guidelines The Assumption of Situational Specificity and Emphasis on Local Validity Uniform Guideline When a test is used to predict At times, sufficient accumulated Sect. 7B Criterion-related the same or similar criteria validity evidence is available for validity studies conducted by (e.g., performance of a given one test user, or described in a selection procedure to justify job) at different times or in test manuals and the its use in a new situation different places, it is typically without conducting a local professional literature, will be found that observed testvalidation research study. considered acceptable for use by another user when the criterion correlations vary In these instances, use of the substantially. In particular following requirements are met. selection procedure may be meta-analytic analyses have based on demonstration of the shown that in some domains. 1. Validity evidence. Evidence generalized validity inferences from that selection procedure much of this variability may be from the available studies due to statistical artifacts such meeting the standards of coupled with a compelling as sampling fluctuations and section 14B below clearly argument for its applicability to demonstrates that the selection variations across validation the current situation. studies in the ranges of test procedure is valid. scores and in the reliability of Although neither mutually 2.Job similarity. The incumbents in the user's job and the incumbents exclusive nor exhaustive, the criterion measures. several strategies for job or group of jobs on which the Thus, statistical summaries of generalizing validity evidence have been delineated: (a) validity study was conducted perform substantially the same major work behaviors, as shown past validation studies in similar situations may be useful in transportability, (b) synthetic estimating test-criterion by appropriate job analyses both on the job or group of jobs on validity/job component validity, relationships in a new and (c) meta-analytic validity generalization. which the validity study was performed and on the job for which the selection procedure is to be used, and



Standards	Principles	Uniform Guidelines
Any significant disparities that might limit the applicability of the meta-analytic findings to the local situation should be noted explicitly.	The rules by which the researchers categorized the work and jobs studied, the selection procedures used, the definitions of what the selection procedure is measuring, the job performance criteria used, and other study characteristics that were hypothesized to impact the study results should be fully reported.  Generalizing validity evidence from meta-analytic results is often more useful than a single study.	
	However, if important conditions in the operational setting are not represented in the meta-analysis (e.g., the local setting involves a managerial job and the meta-analytic data base is limited to entry-level jobs), a local individual study may be more accurate than the average predictor-criterion relationship reported in a meta-analytic study.	

Sou	irce of Validity Evide	nce 🤻 🔭
Standards	Principles	Uniform Guidelines
Evidence Based on Test Content	Content-Related Evidence Evidence Based on Response	There are three concepts which can be used to validate a selection procedure. These
Evidence Based on Response Processes	Processes  Evidence Based on the Internal	concepts reflect different approaches to investigating the job relatedness of selection
Evidence Based on Internal Structure	Structure of the Test  Evidence Based on the	procedures and may be interrelated in practice. They are (1) criterion-related
Evidence Based on Relations to Other Variables	Relationship Between Scores on Predictors and Other Variables	validity, (2) content validity, and (3) construct validity.
Evidence Based on Consequences of Testing	Evidence Based on Consequences of Personnel	Sec. 5. General standards for validity studiesA. Acceptable types of validity studies. For
Thus evidence about	Decisions	the purposes of satisfying these guidelines, users may
consequences is directly relevant to validity when it can be traced to a source of invalidity such as construct underrepresentation or	In recent years, one school of thought has advocated incorporating examination of consequences of the use of	rely upon criterion-related validity studies, content validity studies or construct validity studies, in accordance
construct-irrelevant components. Evidence about consequences	predictors in the determination of validity. This perspective	with the standards set forth in the technical standards of
that cannot be so traced is not directly relevant to the validity of the intended interpretations about	views unintended negative consequences as weakening	these guidelines, section 14 below.

****				
Standards	Principles	Uniform Guidelines	**	
	Although evidence of negative consequences may influence policy or practice decisions concerning the use of predictors, these <i>Principles</i> and the <i>Standards</i> take the view that such evidence is relevant to inferences about validity only if the negative consequences can be attributed to the measurement properties of the selection procedure itself.	New strategies for showing the validity of selection procedures will be evaluated as they become accepted by the psychological profession.	***	



Standards	Principles	Uniform Guidelines	3
Threats to the Fair and Valid Interpretations of Test Scores  A prime threat to fair and valid interpretations of test scores is the presence of construct-irrelevant aspects of the test or testing process that may result in the systematic lowering or raising of scores for identifiable groups of test takers.  Such construct-irrelevant components of scores may be introduced by inappropriate sampling of test content, lack of clarity in test instructions, item complexities that are unrelated to the construct being measured, and/or use of scoring criteria that may favor one group more than another.	Alternatively, if the group difference on the selection procedure is consistent with differences between the groups in the work behavior or performance predicted by the procedure, the finding of group differences could actually support the validity argument.  In this case, negative consequences from test use constitute a policy issue for the user, rather than indicate negative evidence concerning the validity of the selection procedure.	This normally means that it must show a clear relation between performance on the selection procedure and performance on the job.	

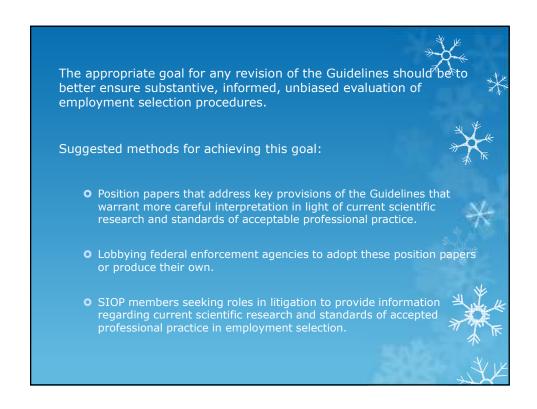
			-31
Standards	Principles	Uniform Guidelines	**
Standard 3.6  Where credible evidence indicates the possibility that test scores may differ in meaning for relevant subgroups in the intended test taker population, the reliability and validity of score interpretations for intended uses for individuals from those subgroups should be examined.  Comment: Subgroup mean differences do not in and of themselves indicate lack of fairness, but such differences should trigger follow up studies to identify the potential causes of such differences. Both construct underrepresentation and sources of construct irrelevant variance should be investigated as potential causes of subgroup differences through quantitative and/or qualitative study.			

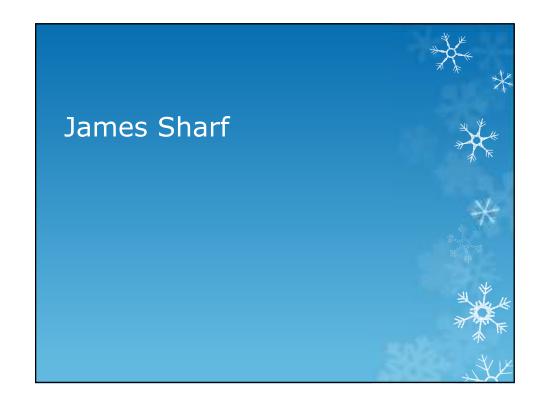
		prediction
Standards	Principles	Uniform Guidelines
Fairness as Lack of Bias The term <i>predictive bias</i> may be used when evidence is found that differences exist in the patterns of associations between test scores and other variables for different groups, bringing with it concerns about bias in the inferences drawn from the use of test scores.	The fourth meaning views fairness as a lack of predictive bias. This perspective views predictor use as fair if a common regression line can be used to describe the predictor-criterion relationship for all subgroups of interest; subgroup differences in regression slopes or intercepts signal predictive bias.	
Differential prediction is examined using regression analysis. One approach examines slope and intercept differences between two targeted groups (e.g., Black vs. White), while another examines systematic deviations from a common regression line for any number of groups of interest.	There is broad scientific agreement on this definition of predictive bias, but there is no similar broad agreement that the lack of predictive bias can be equated with fairness.  Thus, there are multiple perspectives on fairness.	

Standards	Principles	Uniform Guidelines
Standard 3.7  When criterion-related validity evidence is used as a basis for test score interpretations about future performance on some criterion, differential prediction should be evaluated for subgroups for which credible prior evidence or theory suggests the likely existence of such differential prediction, where sample sizes permit.	There is agreement that issues of equitable treatment, predictive bias, and scrutiny for possible bias when subgroup differences are observed, are important concerns in personnel selection; there is not, however, agreement that the term "fairness" can be uniquely defined in terms of any of these issues.	

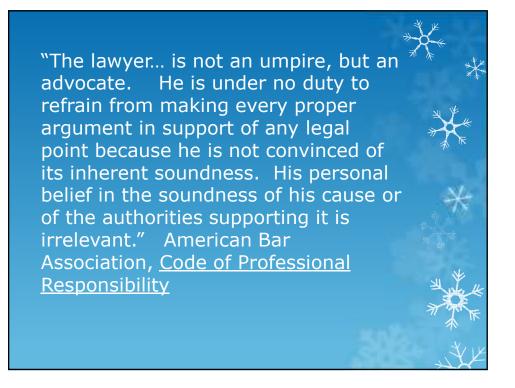
## "The bulk of the Guidelines deals with questions such as those discussed in the above paragraphs. Not all such questions can be answered simply, nor can all problems be addressed in the single document. Once the guidelines are issued, they will have to be interpreted in light of changing factual, legal, and professional circumstances."

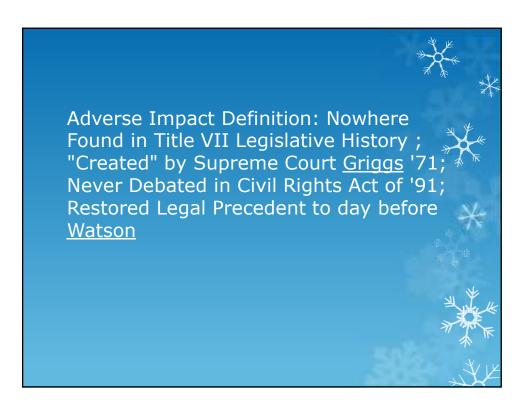








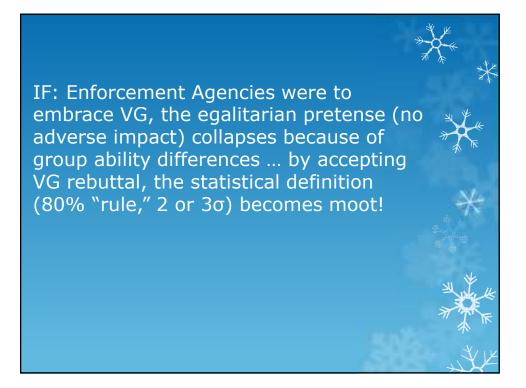




Watson: "The burden of proving a prima facie case is "not onerous," and the employer in turn may rebut it simply by producing some evidence that it had legitimate, non-discriminatory reasons for the decision. ... The ultimate burden of persuading the trier of fact that the defendant intentionally discriminated against the plaintiff remains at all times with the plaintiff."



		***
LEFT	RIGHT	*
Equal Employment	<b>Equal Opportunity</b>	<b>₩</b>
Gov't redistribute wealth/ reduce inequality	Market place: reward talent	****
Group rights	Individual rights	*
Equal results: but for discrimination	Valid decisions	
		NA WAR



Sharf's Bet: The LEFT will not give ground on their **legal** achievement (<u>Griggs</u> statistical "equal results" definition) and their **political** achievement (restoring the <u>Griggs</u> definition in the Civil Rights Act of '91). The Supreme Court created the statistical definition AND the Supreme Court will take it away.

