

FOR PUBLICATION
UNITED STATES COURT OF APPEALS
FOR THE NINTH CIRCUIT

RONALD L. OBREY, JR., <i>Plaintiff-Appellant,</i> v. HANSFORD T. JOHNSON, in his capacity as the Acting Secretary of the Navy, <i>Defendant-Appellee.</i>

No. 03-16849
D.C. No.
CV-02-00033-
MLR/LEK
OPINION

Appeal from the United States District Court
for the District of Hawaii
Manuel L. Real, District Judge, Presiding

Argued and Submitted
November 3, 2004—Honolulu, Hawaii

Filed March 4, 2005

Before: Melvin Brunetti, Susan P. Graber, and Jay S. Bybee,
Circuit Judges.

Opinion by Judge Bybee

COUNSEL

Clayton C. Ikei, Honolulu, Hawaii, for the plaintiff-appellant.

E. Roy Hawkens, Jeffrey Clair, Appellate Staff Civil Division, Department of Justice, Washington, D.C., for the defendant-appellee.

OPINION

BYBEE, Circuit Judge:

This appeal requires us to clarify and apply the harmless error test applicable to civil trials in our circuit.

I.

Appellant, Ronald L. Obrey, Jr., originally filed suit for declaratory and injunctive relief, alleging that he was twice denied a promotion to the position of Production Resource Manager at the Pearl Harbor Naval Shipyard (hereinafter, the “Shipyard”) on the basis of his race in violation of Title VII of the Civil Rights Act of 1964, as amended, 42 U.S.C. § 2000e et seq. (2000). Obrey alleged that the defendant, the Secretary of the Navy, had engaged in a pattern or practice of discriminating against qualified candidates of Asian-Pacific ancestry in favor of Caucasian applicants for senior management positions at the Shipyard. In a pre-trial hearing, the district court issued several evidentiary rulings excluding the principal evidence supporting Obrey’s pattern or practice claim. After a jury trial, judgment was entered against Obrey. The district court’s evidentiary rulings form the basis for this appeal.

The Pearl Harbor Shipyard is one of four Navy shipyards operated by the Navy organizational unit, the Naval Sea Systems Command. Obrey, an Asian-Pacific Islander, has, from 1995-2002, worked as a Project Superintendent at the Shipyard. In 2002, Obrey applied for the Production Resource Manager’s (“PRM”) position at the Shipyard, a position which carried a promotion from his current grade level of GM-14 to a GS-15 grade. Nine other individuals also applied. Pursuant to Navy guidelines, the applicants were rated in three categories, including relevant knowledge, ability to plan and manage resources, and ability to perform supervisory management functions. On the basis of this rating, Obrey was

ranked sixth out of ten applicants during the first round of hiring, and fifth out of the eight competitive applicants in the second round. The PRM position was subsequently offered to Ernest Chamberlain in the first round of hiring, and then David Reilly in the second, both of whom are Caucasian males and both of whom declined the offer. Recruitment was then cancelled.

In this appeal, Obrey claims that the district court abused its discretion in failing to admit three pieces of evidence: (1) a statistical report showing a correlation between race and promotion at the Shipyard; (2) the testimony of a Shipyard employee who recalled conversations in which Shipyard officials expressed discriminatory bias toward the local Asian-Pacific Islanders; and (3) the anecdotal testimony of three Shipyard employees who also believed they had suffered race discrimination at the Shipyard. The Navy argues that the exclusion was proper but that, even if the district court erred, the error was harmless. Addressing each evidentiary ruling in turn, we find that the district court's decision excluding this evidence was an abuse of discretion as to all. We further conclude that the error was not harmless.

A.

The district court denied Obrey's motion in limine to admit statistical evidence regarding hiring practices for senior-level positions at the Shipyard. The hiring practice evidence at issue was compiled through discovery and included the hiring history of the Pearl Harbor Shipyard for the period 1999-2002. Obrey retained James Dannemiller, a statistician with SMS Research & Marketing Services, Inc., to analyze this data and provide a statistical report and opinion. Dannemiller's report concludes that "[t]here is no statistical evidence . . . that the selection process for GS13 through GS15 positions between 1999 and 2002 were unbiased with respect to race."

The government challenged the admission of Dannemiller's report on the ground that it was so incomplete that it was inadmissible as irrelevant, unfairly prejudicial, and unreliable. See FED. R. EVID. 402, 403, 702. In the government's view, the statistical analysis was inadmissible because it failed to account for the relative qualifications of the applicants being studied. The district court denied Obrey's motion to admit Dannemiller's statistical evidence. Although the court did not specify its reasons, presumably its ruling was based on the perceived irrelevance and unreliability of the statistics. While we review evidentiary rulings for an abuse of discretion, *Coursen v. A.H. Robins Co., Inc.*, 764 F.2d 1329, 1333 (9th Cir.), amended by, 773 F.2d 1049 (9th Cir. 1985), neither of these reasons warrants exclusion in this case.

Obrey's claim was premised on the theory that the Navy had engaged in a pattern or practice of discriminatory hiring practices. Employment discrimination claims styled in this manner are governed by "controlling legal principles that are relatively clear." *Int'l Bhd. of Teamsters v. United States*, 431 U.S. 324, 335 (1977). Obrey's theory of discrimination was that the Navy regularly and purposefully treated the local Asian-Pacific Islanders less favorably than white persons by refusing to promote minority group members on an equal basis. His suit thus raised as factual issues "whether there was a pattern or practice of such disparate treatment and, if so, whether the differences were 'racially premised.'" *Id.* at 335 (quoting *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 805 n.18 (1973)).

As the plaintiff, Obrey bore the initial burden of making out a prima facie case of discrimination. *Cooper v. Fed. Reserve Bank of Richmond*, 467 U.S. 867, 874 (1984). And, because he alleged a systemwide pattern or practice of resistance to the full enjoyment of Title VII rights, Obrey ultimately had to prove "more than the mere occurrence of isolated or 'accidental' or sporadic discriminatory acts." *Teamsters*, 431 U.S. at 336. He had to establish, by a prepon-

derance of the evidence, that racial discrimination was the Navy's "standard operating procedure—the regular rather than the unusual practice." *Id.* By "demonstrating the existence of a discriminatory pattern or practice," Obrey would "establish[] a presumption that [he] had been discriminated against on account of race." *Cooper*, 467 U.S. at 875 (citing *Franks v. Bowman Transp. Co.*, 424 U.S. 747, 772 (1976)).

[1] In a case in which the plaintiff has alleged that his employer has engaged in a "pattern or practice" of discrimination, "[s]tatistical data is relevant because it can be used to establish a general discriminatory pattern in an employer's hiring or promotion practices. Such a discriminatory pattern is probative of motive and can therefore create an inference of discriminatory intent with respect to the individual employment decision at issue." *Diaz v. Am. Tel. & Tel.*, 752 F.2d 1356, 1363 (9th Cir. 1985);¹ *see also McDonnell Douglas*, 411 U.S. at 805 n.19 ("The District Court may, for example, determine, after reasonable discovery that the (racial) composition of defendant's labor force is itself reflective of restrictive or exclusionary practices.") (internal quotation marks omitted); *Coral Constr. Co. v. King County*, 941 F.2d 910, 918 (9th Cir. 1991) ("[F]or purposes of Title VII, '[w]here gross statistical disparities can be shown, they alone may in a proper case constitute prima facie proof of a pattern or practice of discrimination.'") (quoting *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 307-08 (1977)); *Diaz*, 752 F.2d at 1363 ("In some cases, statistical evidence alone may be sufficient to establish a *prima facie* case. . . . Even when not sufficient to establish a *prima facie* case, statistical evidence is helpful in showing that an employer's articulated reason for the employment decision is pretextual . . .") (citations omitted).²

¹The Navy argues that Obrey abandoned his pattern or practice claim at trial. If Obrey did so, it was because the trial court excluded his evidence. Any abandonment was compelled and was not a waiver of the claim.

²The Supreme Court has suggested on several occasions that a statistical comparison is a valuable tool with which to evaluate a claim of employ-

Obrey's statistical evidence was not rendered irrelevant under Rule 402 simply because it failed to account for the relative qualifications of the applicant pool. *See* FED. R. EVID. 402 ("All relevant evidence is admissible, except as otherwise provided [by law]. Evidence which is not relevant is not admissible.") A statistical study may fall short of proving the plaintiff's case, but still remain relevant to the issues in dispute. The Dannemiller study may be relevant, and therefore admissible, even if it is not sufficient to establish Obrey's *prima facie* case or a claim of pretext. Thus, objections to a study's completeness generally go to "the weight, not the admissibility of the statistical evidence," *Mangold v. Cal. Pub. Utils. Comm'n*, 67 F.3d 1470, 1476 (9th Cir. 1995), and should be addressed by rebuttal, not exclusion, *Teamsters*, 431 U.S. at 340. As the Court has pointed out,

Statistics showing racial or ethnic imbalance are probative . . . because such imbalance is often a telltale sign of purposeful discrimination; Considerations such as small sample size may, of course, detract from the value of such evidence, and evidence showing that the figures for the general population might not accurately reflect the pool of qualified job applicants would also be relevant.

Teamsters, 431 U.S. at 339-40 n.20 (citations omitted); *see also Bazemore v. Friday*, 478 U.S. 385, 400 (1986) (per curiam) ("Normally, failure to include variables will affect the analysis' probativeness, not its admissibility.") (Brennan, J., concurring in part); *Hemmings v. Tidyman's Inc.*, 285 F.3d 1174, 1188-89 (9th Cir. 2002) ("[T]he law does not require

ment discrimination. *See, e.g., Furnco Constr. Corp. v. Waters*, 438 U.S. 567, 580 (1978) (district court entitled to consider the racial mix of the workforce); *Teamsters*, 431 U.S. at 339 ("[O]ur cases make it unmistakably clear that statistical analyses have served and will continue to serve an important role in cases in which the existence of discrimination is a disputed issue.") (internal quotation marks omitted).

the near-impossible standard of eliminating *all* possible non-discriminatory factors. . . . We cannot say that the exclusion of preferences, individual qualifications, and education rendered the data set so incomplete ‘as to be irrelevant.’ ” (quoting *Bazemore*, 478 U.S. at 400) (emphasis in original); *cert. denied*, 537 U.S. 1110 (2003); *Maitland v. Univ. of Minn.*, 155 F.3d 1013, 1017 (8th Cir. 1998) (“[A] regression analysis does not become inadmissible as evidence simply because it does not include every variable that is quantifiable and may be relevant to the question presented. . . . [I]t is for the finder of fact to consider the variables that have been left out of an analysis, and the reasons given for the omissions, and then to determine the weight to accord the study’s results”); *Wilmington v. J.I. Case Co.*, 793 F.2d 909, 920 (8th Cir. 1986) (“Virtually all the inadequacies in the expert’s testimony urged here by [the defendant] were brought out forcefully at trial These matters go to the weight of the expert’s testimony rather than to its admissibility.”).

In some cases, statistical evidence may suffer from serious methodological flaws and can be excluded, consistent with the trial court’s “gatekeeping” power, under Rule 702. See *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 156-57 (1999); *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 589-90 (1993).³ Factors which may bear on admissibility include: (1) whether the “scientific knowledge . . . can be (and has been) tested”; (2) whether “the theory or technique has been sub-

³Rule 702 provides:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

FED. R. EVID. 702.

jected to peer review and publication”; (3) “the known or potential rate of error”; and (4) “general acceptance.” *Daubert*, 509 U.S. at 593-94. The Rule 702 inquiry is a “flexible one” whose “overarching subject is the scientific validity and thus the evidentiary relevance and reliability [] of the principles that underlie a proposed submission.” *Id.* at 594-95; *see also Kumho Tire*, 526 U.S. at 141 (“[T]he test of reliability is ‘flexible,’ and *Daubert*’s list of specific factors neither necessarily nor exclusively applies to all experts or in every case.”).

[2] Here, the Dannemiller study is based entirely on statistical disparities. While we, and other courts, have commented on the inadequacy of such studies, we have typically done so in the context of finding insufficient evidence to support a *prima facie* case of discrimination, and not to rule those studies inadmissible for purposes of Rule 702. *See, e.g., Coleman v. Quaker Oats Co.*, 232 F.3d 1271, 1283 (9th Cir. 2000) (“Because [the statistics] fail to account for many factors pertinent to [the plaintiff], we conclude that the statistics are not enough to take this case to trial.”); *Ottaviani v. State Univ. of N.Y. at New Paltz*, 875 F.2d 365, 370-75 (2d Cir. 1989) (statistical evidence was not “statistically significant” enough to establish a *prima facie* case of discrimination); *Gay v. Waiters’ & Dairy Lunchmen’s Union, Local No. 30*, 694 F.2d 531, 553 (9th Cir. 1982) (“[S]tatistical evidence, standing alone, was insufficient to establish a *prima facie* case.”). As a general matter, so long as the evidence is relevant and the methods employed are sound, neither the usefulness nor the strength of statistical proof determines admissibility under Rule 702. *See Metabolife Int’l, Inc. v. Wornick*, 264 F.3d 832, 843 (9th Cir. 2001) (“Rather than disqualify the study because of ‘incompleteness’ . . . , the district court should examine the soundness of the methodology employed.”).

[3] In sum, Dannemiller’s study was relevant for what it purported to analyze: the race of managers selected at the Shipyard compared to the race of those who applied for mana-

gerial positions. While, by itself, this cannot constitute proof that the Navy discriminated against Obrey, *see Cooper*, 467 U.S. at 876, it should have been admitted for whatever probative value it had. Since the defendant's objections to the admission of Dannemiller's study went to weight and sufficiency rather than admissibility, we conclude that the district court abused its discretion when it excluded this evidence.

B.

The district court also excluded the testimony of a single Shipyard worker, Mr. Toyama, on the grounds that his evidence was irrelevant. FED. R. EVID. 401 (“‘Relevant evidence’ means evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.”). Toyama was expected to testify that Shipyard officials had informed him that off-yard employees were rotated to Pearl Harbor on a temporary basis because the “local” workers “were not good enough” and “can’t do a good job.”

[4] Toyama’s testimony was plainly relevant to the issue of whether the defendant preferred off-yard, predominantly Caucasian, workers over the “local” Asian-Pacific Islanders. We have observed that “evidence that the defendant has made disparaging remarks about the class of persons to which plaintiff belongs[] may be introduced to show that the defendant harbors prejudice toward that group.” *Lam v. Univ. of Haw.*, 164 F.3d 1186, 1188 (9th Cir. 1999) (internal quotation marks omitted). It tends to show “a defendant’s discriminatory state of mind.” *Id.*⁴

⁴The Navy argues that these comments were not directed at the “locals” —meaning the Asian-Pacific Islanders—but were critical of the general efforts of all Navy employees at the Shipyard. The inferences to be drawn from these comments should be resolved by a jury.

[5] Toyama's testimony was also relevant to whether the Navy's proffered race-neutral reasons for preferring off-yard workers was a pretext for unlawful race discrimination. Obrey asserts that Toyama also would have challenged the Navy's claim that off-yard managers were more capable of performing their tasks within the Shipyard's budget by demonstrating that the imported managers were funded by budgeted funds separate and apart from the Shipyard's budget. According to Obrey, this testimony would have cast doubt on the Navy's explanation by demonstrating that the off-yard managers exerted no effect whatsoever on the Shipyard's budget.

[6] Because Toyama's testimony tended to make the existence of discriminatory bias and pretext more probable than it would be without his testimony, we find that the district court abused its discretion by excluding this evidence.

C.

The district court also excluded the testimony of three Shipyard workers, Kawachi, Pestana and Tai See, who were prepared to testify that the Shipyard discriminated against them on the basis of race when it failed to select them for supervisory positions. The court found that the testimony at issue would require the jury to assess the discrimination claims of each of the three proposed witnesses by, essentially, conducting three abbreviated employment discrimination trials. The court concluded that the testimony should be excluded on the basis of Federal Rule of Evidence 403, presumably because considerations of undue delay and waste of time outweighed its probative value. *See* FED. R. EVID. 403 ("Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.").

[7] Like statistical evidence, anecdotal evidence of past discrimination can be used to establish a general discriminatory

pattern in an employer's hiring or promotion practices. While such evidence might prove inadmissible in the typical case of individual discrimination, in a case involving a claim of discriminatory pattern or practice "the combination of convincing anecdotal and statistical evidence is potent." *Coral Constr. Co.*, 941 F.2d at 919. It is commonplace that a plaintiff attempting to establish a pattern or practice of discriminatory employment will present some anecdotal testimony regarding past discriminatory acts. *See, e.g., Rossini v. Ogilvy & Mather, Inc.*, 798 F.2d 590, 604 (2d Cir. 1986) ("In evaluating all of the evidence in a discrimination case, a district court may properly consider the quality of any anecdotal evidence or the absence of such evidence."); *Coates v. Johnson & Johnson*, 756 F.2d 524, 532 (7th Cir. 1985) ("The plaintiffs' prima facie case will thus usually consist of statistical evidence demonstrating substantial disparities in the application of employment actions as to minorities and the unprotected group, buttressed by evidence of . . . specific instances of discrimination."); *Valentino v. United States Postal Serv.*, 674 F.2d 56, 69 (D.C. Cir. 1982) ("[W]hen the statistical evidence does not adequately account for the diverse and specialized qualifications necessary for (the positions in question), strong evidence of individual instances of discrimination becomes vital to the plaintiff's case.") (internal quotation marks omitted); *Garcia v. Rush-Presbyterian-St. Luke's Med. Ctr.*, 660 F.2d 1217, 1225 (7th Cir. 1981) ("We find very damaging to plaintiffs' position the fact that not only was their statistical evidence insufficient, but that they failed completely to come forward with any direct or anecdotal evidence of discriminatory employment practices by defendants. Plaintiffs did not present in evidence even one specific instance of discrimination.").

[8] We recognize, however, that the district court retains broad discretion to determine whether the probative value of the evidence at issue is substantially outweighed by considerations of "undue delay, waste of time, or needless presentation of cumulative evidence." FED. R. EVID. 403; *see also R.B.*

Matthews, Inc. v. Transamerica Transp. Servs., Inc., 945 F.2d 269, 272 (9th Cir. 1991) (“Trial judges have wide discretion to exclude evidence given their presence at the trial and because the considerations arising under Rule 403 are ‘susceptible only to case-by-case determinations, requiring examination of the surrounding facts, circumstances, and issues.’”) (quoting *United States v. Layton*, 767 F.2d 549, 554 (9th Cir. 1985)). Nevertheless, none of the testimony that the appellant attempted to offer into evidence so clearly involved delay that was “undue” or a “waste of time” or was cumulative of other evidence that it was excludable. Rather, the testimony was offered to show that the defendant had a discriminatory motive when it denied his promotion because it had unlawfully rejected other applicants in circumstances similar to his, and tended to support his pattern or practice theory. While the jury naturally has to determine the credibility of witness testimony in order to assess the weight it should be accorded, this is not the sort of undue delay and waste of time that the Rules contemplate.

[9] We acknowledge that the trial court was properly concerned with the prospect of mini-trials on the witnesses’ own claims of discrimination. The trial court should have first addressed these concerns with the parties through other, less restrictive means. On balance, we believe that this proposed testimony was likely to be relevant, and Rule 403 considerations do not warrant exclusion in this case. Consequently, we find that the district court abused its discretion when it excluded this testimony. On remand, the district court, of course, will retain discretion to decide that the witnesses’ claims so overwhelm the issues in the trial that their testimony must be excluded under Rule 403.

II.

[10] Turning to the question of harmless error, we note, initially, that judicial error alone does not mandate reversal. Rather, in order to reverse, we must find that the error

affected the substantial rights of the appellant. *See* FED. R. EVID. 103(a) (“Error may not be predicated upon a ruling which admits or excludes evidence unless a substantial right of the party is affected”); FED. R. CIV. P. 61 (“The court at every stage of the proceeding must disregard any error or defect in the proceeding which does not affect the substantial rights of the parties.”). In other words, we require a finding of prejudice. *See Kisor v. Johns-Manville Corp.*, 783 F.2d 1337, 1340 (9th Cir. 1986). Although frequently termed a “harmless error” analysis, this inquiry turns on the distinction between the burden of proof required in civil and criminal trials: “Just as the verdict in a civil case need only be more probably than not true, so an error in a civil trial need only be more probably than not harmless.” *Haddad v. Lockheed Cal. Corp.*, 720 F.2d 1454, 1459 (9th Cir. 1983).

[11] In a somewhat contradictory fashion, however, we have formulated two variations of the test for prejudice in civil cases. In *Haddad*, we held that the reviewing court *must* find prejudice *unless* it concludes that the verdict is “more probably than not untainted by the error.” *Id.* Purporting to restate the standard set forth in *Haddad*, we later wrote in *Kisor*, that “[t]o reverse, we must say that more probably than not, the error tainted the verdict.” *Kisor*, 783 F.2d at 1340 (citing *Haddad*, 720 F.2d at 1459). As we noted in *Pau v. Yosemite Park & Curry Co.*, 928 F.2d 880, 888 & n.2 (9th Cir. 1991), and *Ortega v. O’Connor*, 50 F.3d 778, 780 n.2 (9th Cir. 1995), this restatement effected more than a mere semantic change. Rather, “in a close case, where the reviewing court is uncertain of the effect of an evidentiary error on the jury’s verdict, these two standards create contradictory presumptions.” *Id.* Under *Haddad*’s formulation, we presume prejudice; under *Kisor*, we appear to presume the opposite. *Pau*, 928 F.2d at 888 n.2.

Making matters worse, we have inconsistently applied *Haddad* and *Kisor*. We have cited both without recognizing the contradiction. *See, e.g., Baker v. Delta Air Lines, Inc.*, 6

F.3d 632, 639 (9th Cir. 1993) (quoting *Kisor* and noting its reliance on *Haddad*); *Cassino v. Reichhold Chems., Inc.*, 817 F.2d 1338, 1345 (9th Cir. 1987). We have applied one or the other without recognizing or purporting to resolve the contradiction. See, e.g., *Blind-Doan v. Sanders*, 291 F.3d 1079, 1082 (9th Cir. 2002) (restating the standard a la *Kisor* and citing *Pau*); *Tennison v. Circus Circus Enters.*, 244 F.3d 684, 688 (9th Cir. 2001) (same); *Beachy v. Boise Cascade Corp.*, 191 F.3d 1010, 1015-16 (9th Cir. 1999) (quoting *Haddad*); *Oliver v. United States*, 921 F.2d 916, 920 (9th Cir. 1990) (quoting *Haddad*); *Brown v. Sierra Nev. Mem'l Miners Hosp.*, 849 F.2d 1186, 1190 (9th Cir. 1988) (quoting *Kisor*). And we have recognized the contradiction but declined to address it because the case before us was not so close that the presumption would affect the outcome. See, e.g., *Ortega*, 50 F.3d at 780 n.2; *Ackley v. W. Conference of Teamsters*, 958 F.2d 1463, 1470 & n.4 (9th Cir. 1992); *Pau*, 928 F.2d at 888 & n.2. Because this appeal presents precisely such a close case, we find it necessary to resolve this conflict.

[12] We must follow *Haddad*. We believe that our contrary language in *Kisor* inadvertently reversed the presumption of prejudice observed in *Haddad*. See *Pau*, 928 F.2d at 888 (characterizing *Kisor*'s language as an "inadvertent misstatement"). Cf. *Coursen v. A.H. Robins Co.*, 764 F.2d at 1334, 1337, 1338, 1340, *amended by*, 773 F.2d 1049 (9th Cir. 1985) (before *Kisor*, repeatedly citing *Haddad* but inconsistently restating the *Haddad* standard — twice correctly, and twice inadvertently reversing the presumption a la *Kisor*). Nothing in *Kisor* suggested that intervening Supreme Court or en banc decisions or new rules had rendered *Haddad*'s holding incorrect or amenable to reinterpretation, or that we intended to actually reinterpret *Haddad*. Rather, *Kisor*'s citation of *Haddad* without any additional commentary indicates our intent to remain faithful to *Haddad*. Cf. *O'Neal v. McAninch*, 513 U.S. 432, 438-39 (1995) (stating that language in a prior opinion that suggested a reversal of the burden of proof for harmless error was "not determinative" because the restatement was

inconsistent with the Court's intention in that opinion to merely apply precedent). Moreover, even if the *Haddad* standard were open to revisitation by a three-judge panel, addressing it in *Kisor* would have been inappropriate for the same reason we declined to address the *Haddad-Kisor* conflict in *Ortega, Ackley, and Pau*: The presumption was irrelevant because it was not a "close case." See *Kisor*, 783 F.2d at 1342 (finding that "the verdict was probably tainted" and reversing, even while purporting to presume harmlessness). We therefore decline to recognize *Kisor* as affecting prior precedent as to the precise formulation of the harmless error standard.

Apart from its precedential pedigree, we adopt *Haddad's* formulation of the harmless error standard for the additional reason that we believe it to be correct on the merits. First, *Haddad* is in keeping with "the original common-law harmless-error rule [that] put the burden on the beneficiary of the error either to prove that there was no injury or to suffer a reversal of his erroneously obtained judgment." *Chapman v. California*, 386 U.S. 18, 24 (1967).

Second, we recognized in *Haddad* that "appellate courts have three possible standards of review: harmless beyond a reasonable doubt; high probability of harmlessness; and more probably than not harmless." 720 F.2d at 1458 n.7 (citing ROGER TRAYNOR, *THE RIDDLE OF HARMLESS ERROR* (1972)); see also *Neder v. United States*, 527 U.S. 1, 7 (1999) (noting that in criminal cases constitutional errors affecting substantial rights require automatic reversal, and all other constitutional errors are disregarded only if harmless beyond a reasonable doubt); *United States v. Valle-Valdez*, 554 F.2d 911, 915-16 (9th Cir. 1977) (recognizing the same three possible standards and applying the more-probable-than-not standard to nonconstitutional errors in criminal cases). Each of these "possible" formulations implies a presumption of prejudice; none presumes harmlessness.

[13] Third, presuming prejudice, rather than harmlessness, is required by Supreme Court precedent. In *O'Neal*, the Court

rejected both the premise and conclusion of the argument that a presumption of harmlessness applies in civil cases and that therefore such a presumption should apply in habeas cases. The Court held:

[P]recedent suggests that civil and criminal harmless-error standards do not differ in their treatment of grave doubts as to the harmlessness of errors affecting substantial rights. . . . [E]ven if, for argument's sake, we were to assume that the civil standard for judging harmlessness applies to habeas proceedings (despite the fact that they review errors in state *criminal* trials), it would make no difference with respect to the matter before us. For relevant authority rather clearly indicates that, either way, the courts should treat similarly the matter of "grave doubt" regarding the harmlessness of errors affecting substantial rights, and as *Kotteakos* provides.

O'Neal, 513 U.S. at 441-42 (referring to *Kotteakos v. United States*, 328 U.S. 750 (1946)). *Kotteakos* provides that "[if] the error itself had substantial influence . . . or if one is left in grave doubt [i.e., equipoise], the conviction cannot stand." 328 U.S. at 764-65. Thus, the harmless error standard we apply in civil cases must be consistent with the standard we apply to nonconstitutional errors in criminal cases: "we must reverse . . . unless it is more probable than not that the error did not materially affect the verdict." *United States v. Morales*, 108 F.3d 1031, 1040 (9th Cir. 1997) (en banc). The party benefitting from the error has the burden of persuasion, and "in cases of 'equipoise,' we reverse." *United States v. Seschillie*, 310 F.3d 1208, 1214-15 (9th Cir. 2002). This standard is substantively identical to the standard we applied in *Haddad*, 720 F.2d at 1549, and it is clear from *O'Neal* that we were correct in adopting it for civil cases.

Thus, when reviewing the effect of erroneous evidentiary rulings, we will begin with a presumption of prejudice. That

presumption can be rebutted by a showing that it is more probable than not that the jury would have reached the same verdict even if the evidence had been admitted. *Haddad*, 720 F.2d at 1459.

Applying this standard to the facts before us, the Navy would have us hold that it is more probable than not that the district court's erroneous exclusion of evidence probative of its alleged discriminatory bias and pretext did not taint the jury's verdict. Although recognizing the burden that an additional trial would place on the parties, we decline to do so.

As we noted in *Haddad*: "The danger of the harmless error doctrine is that an appellate court may usurp the jury's function, by merely deleting improper evidence from the record and assessing the sufficiency of the evidence to support the verdict below." 720 F.2d at 1459 (citing *Kotteakos*, 328 U.S. at 764-65; TRAYNOR, *THE RIDDLE OF HARMLESS ERROR* at 18-22). While this danger has less practical importance where the litigant merely has a right to a jury verdict that "more probably than not" corresponds to the truth, our task on appeal remains meaningful: We must determine whether the evidentiary error of which appellant complains has deprived him of the degree of certainty to which he is entitled. *Haddad*, 720 F.2d at 1459.

[14] We cannot conclude, based upon the facts of this case, that the erroneous exclusion of evidence directly probative of the defendant's discriminatory bias and pretext did not taint the jury's verdict. The evidence at issue was not merely tangential or cumulative; rather, it was directly probative of the central issues in dispute. Although the Dannemiller study is in the record, neither Toyama nor the three Shipyard workers actually testified; we know only what Obrey claimed they would say. We are reluctant to judge a fact-intensive case on the basis of mere proffers of evidence. We thus cannot state that it is more probable than not that the jury was unaffected by the erroneous exclusion of the plaintiff's principal evi-

dence. Accordingly, we hold that the district court's erroneous exclusion of the Dannemiller study, the testimony of Mr. Toyama, and the anecdotal testimony of three Shipyard workers was an abuse of discretion requiring reversal. The erroneous exclusion was not harmless.

III.

For the foregoing reasons, the judgment of the district court is REVERSED and the case is REMANDED for proceedings consistent with this opinion.