

«Grados de libertad» y el estudio de casos

Donald T. Campbell

Universidad de Siracusa

Metodología cuantitativa versus cualitativa. La controversia entre los distintos modos de conocimiento «cualitativo/cuantitativo», entre «geisteswissenschaftlich» y «naturwissenschaftlich», entre conocimiento humanista y científico es característica de la mayoría de las ciencias sociales hoy en Estados Unidos. En campos como la Sociología y la Psicología Social, muchos de los más capaces y dedicados de nuestros estudiantes graduados están optando, cada vez más, por el método cualitativo/humanístico. En la Ciencia Política hay una continua división entre estas líneas. Solamente la Economía y la Geografía parecen relativamente inmunes a esta dicotomía.

Inevitablemente, esta escisión ha sobrepasado la investigación evaluativa tomando la forma de una controversia sobre la legitimidad del paradigma cuantitativo-experimental para la evaluación de programas (e. g. Weiss & Rein, 1969, 1970; Guttentag, 1971, 1973; Campbell, 1970; Salasin, 1973). La cuestión no ha sido, para ser exactos, planteada en estos términos. Los críticos, desde lo que yo estoy denominando la posición humanística, están a menudo bien entrenados en métodos cuantitativos-experimentales. Sus críticas específicas están con frecuencia bien fundadas en el propio marco experimentalista: los experimentos de un sólo tratamiento y en una sola situación son profundamente ambiguos como para saber qué es lo que causa qué; existe una precaria rigidez en el sistema de medición, limitando los resultados registrados a aquellas dimensiones anticipadas de antemano; los procesos son descuidados frecuentemente en un programa experimental enfocado sobre el efecto

total de un tratamiento complejo y, en este sentido, conocer tales efectos presenta solamente implicaciones equivocadas para la réplica del programa o para su mejoramiento. Los programas de amplio alcance son a menudo irremediablemente ambiguos en lo que respecta a objetivos e indicadores relevantes; los cambios en el programa de tratamiento durante el curso de un experimento de mejoramiento, aunque sean pragmáticamente indispensables, hacen ininterpretables las comparaciones de las salidas y entradas experimentales; los programas sociales son a menudo realizados pobremente desde el punto de vista del diseño experimental; incluso en situaciones bien controladas, la experimentación es un proceso profundamente tedioso y equivocado; la experimentación es demasiado lenta para ser utilizada políticamente, etc. Todas estas son verdades más que suficientes para motivar una búsqueda intensa de alternativas. Hasta ahora, las alternativas sugeridas desde el método cualitativo (e. g. Weiss & Rein, 1969, 1970; Guttentag, 1971, 1973) no me han convencido. De hecho creo que la observación naturalista de los hechos es un campo intrínsecamente erróneo para la inferencia causal, sea mediante métodos cualitativos o cuantitativos, a causa de la confusión inevitable entre selección y tratamiento. Cualquier esfuerzo para reducir esta confusión tendrá el efecto de hacer las condiciones más «experimentales». Los «experimentos» son, de hecho, simplemente aquel tipo de situación observacional manipulada de una forma óptima para llegar a inferencias causales. Los problemas de inferencia existentes son intrínsecos a los contextos de estos programas dentro de procesos

sociales en marcha. Los diseños experimentales no plantean estos problemas y, de hecho, los aminoran, aunque a veces sólo ligeramente.

En tales objeciones, parece estar implícitas la exigencia de sustitución del indirecto y presuntivo proceso de la ciencia por la clarividencia cualitativa. Pero aunque deba rechazar este aspecto de la objeción humanística, existen otros aspectos de ella, que han motivado estas críticas, a las cuales puedo unir me enteramente. Esta otras críticas pueden ser tituladas «olvido de la evidencia contextual cualitativa y relevante» o «basarse excesivamente en unas cuantas abstracciones cuantificadas despreciando la evidencia cualitativa contraria y suplementaria».

Muchas veces los científicos sociales cuantitativos, bajo la influencia de los apóstoles del positivismo lógico, dan por supuesto que en la Ciencia verdadera el conocimiento cuantitativo sustituye al cualitativo, al conocimiento del sentido común. La situación es, de hecho, bastante diferente. Más exactamente, la ciencia depende de lo cualitativo, del conocimiento del sentido común, aunque en el mejor de los casos vaya más allá de él. La ciencia, al final, contradice algunos aspectos del sentido común, pero lo hace sólo confiando en la mayor parte del sentido común. Tal revisión del sentido común por la ciencia es similar a la revisión del sentido común por el propio sentido común la cual, paradójicamente, sólo puede hacerse confiando más en el sentido común. Consideremos, como un ejemplo, la ilustración de Muller-Lyer (Figura 1).

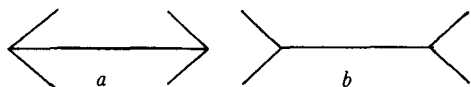


Figura 1

Si se pregunta a un miembro normal de una cultura «carpintera» (Segall, et al., 1966) qué línea es más larga, «a» o «b», responderá que «b». Si se le dá una regla, o se le permite utilizar el filo de una hoja de papel como regla, se convencerá por sí mismo de que está equivocado y que la línea «a» es más larga. Al hacer esta decisión habrá rechazado como inadecuado un producto de la percepción visual al confiar en mayor medida en otras percepciones visuales. Habrá hecho también muchas presunciones, no explícitas en su mayoría, incluyendo la asunción de que la longitud de las líneas ha permanecido relativa-

mente constante durante el proceso de medición, que la regla era más rígida que elástica, que el calor y la humedad de su mano no habían cambiado la longitud de la regla de tal forma que produjera la diferencia de medidas, expandiéndola al aproximarla a la línea «b» y contrayéndola al aproximarse a la línea «a», etc.

Consideremos otro ejemplo, el de un documento científico que contiene teoría y resultados experimentales que demuestran la especial naturaleza de la luz, en dramático contraste con el conocimiento de sentido común. O un documento científico que demuestra que lo que la percepción ordinaria considera «sólidos» son, de hecho, vigas abiertas. Si semejante documento se limitara asimismo a símbolos matemáticos y términos puramente científicos omitiendo el lenguaje ordinario, fracasaría al no permitir una comunicación con otro científico que le permita replicar el experimento y verificar las observaciones. En vez de ello, los escasos términos científicos se introducen en un discurso de lenguaje ordinario precientífico que se presume que el lector entiende. Y en el trabajo de laboratorio original se utiliza y se confía en un lenguaje precientífico de sentido común y en la percepción de objetos, sólidos y luz al extraer las conclusiones que mediante este procedimiento permiten revisar el conocimiento ordinario. En este caso, hay que confiar en el entendimiento del sentido común general para poner a prueba y corregir el entendimiento del sentido común específico.

Relacionado con esto, está el énfasis epistemológico en la identificación de pautas cualitativas como anterior a una identificación de partículas atómicas cuantificables, lo que es opuesto a la intuición del atomista lógico, todavía demasiado difundida (Campbell, 1966). Tal epistemología da más preponderancia a la falibilidad que a la clarividencia, haciendo hincapié en la presumible tendencia al error de tal pauta de identificación más que en la percepción como terreno fiable de certeza. Pero también reconoce que el único camino es la percepción ordinaria, falible, intuitiva y presuntiva. Esto no significa que las percepciones no puedan ser criticadas (Campbell, 1-69) sino que sólo son criticables, como hemos visto, confiando en otras muchas percepciones del mismo nivel epistemológico.

Si aplicamos tal epistemología a la investigación evaluativa, vemos que legitiman inmediatamente la parte de «historia narrativa» de la

mayoría de los informes y sugiere que a esta actividad le sea dado reconocimiento formal en el planteamiento y ejecución del estudio en vez de considerarla sólo como una reflexión a posteriori. Sin esto, los estudios de evaluación no pueden ser interpretados y con más de esto la mayoría serían mejor interpretados. Para que este contenido no sea subjetivo y culpable de sesgos de perspectiva tendríamos que seleccionar mejor a aquellos que registran los hechos y preparar procedimientos formales por los cuales todos los participantes interesados puedan ofrecer adiciones y correcciones a la historia oficial. Tendría que tomarse en consideración la utilización de historiadores entrenados profesionalmente, antropólogos y sociólogos cualitativos. La historia narrativa es parte indispensable del informe final y deberían utilizarse los mejores métodos cualitativos para prepararla.

Deberíamos también reconocer que, durante siglos, participantes y observadores han estado evaluando programas sin beneficiarse de la cuantificación o del método científico. Es este el conocimiento de sentido común en el que nuestra evidencia científica debería apoyarse para sobrepasarlo y no para reemplazarlo. Pero normalmente se le descuida en las evaluaciones cuantitativas, a no ser que sean incluidas unas cuantas anécdotas justificativas, casualmente reunidas. Según la epistemología que defiende, se deberíamos intentar reunir sistemáticamente todas las críticas cualitativas y de sentido común a los programas y las evaluaciones que se hayan generado entre el personal del programa, los clientes del programa y sus familiares, y los observadores de la comunidad. Aunque en esta fase suelen introducirse procedimientos cuantitativos, en la recogida y resumen de los datos tendrían también que emplearse los métodos no cuantitativos de recogida y compilación de datos, (por ejemplo, discusiones de grupo jerárquicamente organizadas). Cuando estas evaluaciones sean contrarias a los resultados cuantitativos, éstos deberían ser considerados como sospechosos hasta que las razones de la discrepancia se hayan aclarado. Ninguno es infalible, evidentemente. Pero para muchos de nosotros, es necesario insistir en que los resultados cuantitativos pueden ser tan erróneos como los cualitativos. Después de todo, en los laboratorios de física, los registros funcionan a veces mal y es generalmente el conocimiento cualitativo, sumado a las presunciones sobre lo que el registro debería mostrar, el que descubre

el error. (Esta idea se encuentra a años luz del mito de que los datos registrados definen operacionalmente los parámetros teóricos).

Escribo con pena que en los Estados Unidos las evaluaciones de programas, (esta sensible utilización de la unión de los dos modos de conocimiento) no se practican todavía. En vez de ello parece optarse por uno u otro modelo. Allí donde, (como en el Modelo de evaluaciones de Ciudades) se han utilizado antropólogos como observadores han participado sustituyendo a los indicadores cuantitativos, los pretest, postest y comparaciones de grupos de control, en vez de completarlos. Un ejemplo en curso de la utilización de los antropólogos se encuentra en el programa de «las Escuelas Experimentales» que comenzó en el Ministerio de Educación y continúa ahora en el Instituto de Educación. En este programa, se fomenta la iniciativa en el sistema escolar y los programas ganadores reciben incrementos substanciales de sus presupuestos (alrededor del 25 %) por llevar a cabo tales innovaciones. Para evaluar algunos de estos programas se ha permitido la contratación, a precio muy alto, de evaluaciones antropológicas de programas únicos. En uno de los casos se implicó a un equipo de cinco antropólogos por un período de cinco años, para estudiar el sistema de enseñanza de una única ciudad de 100.000 habitantes. Los antropólogos carecían de experiencia anterior con otro sistema escolar en Estados Unidos. No se les permitió ningún período de estudio previo antes de que el programa hubiera comenzado (de hecho, llegaron cuando ya se había comenzado el programa). No se preveió que estudiaran otro sistema de enseñanza comparable pero no sujeto a dicho cambio. Creer que, en estas condiciones de observación nada ventajosas, estos observadores cualitativos pueden inferir qué aspectos de los procesos que ellos observan son debidos al nuevo programa innovador, requiere más fe de la que yo tengo, aunque no opinaré hasta ver los resultados. Además, el énfasis del estudio está en las primeras observaciones de los antropólogos mismos, más que en el papel de utilizar a los participantes como informadores. Como resultado, probablemente se dañaran a un lado las observaciones de otros observadores cualitativos mejor situados que los antropólogos. Incluyo aquí a los padres que han tenido otros niños en la escuela antes de producirse el cambio; los profesores que se han incorporado al sistema con experiencia en sistemas

comparables; los profesores que han visto este sistema antes del cambio, durante él y después de él cambio, y los propios estudiantes. Tales observaciones se podrían recoger en masa en forma de cuestionarios. Si es así, sería de desear que se hubieran hecho las preguntas adecuadas antes del programa experimental y en ambas ocasiones en algún sistema de enseñanza comparable, en el que no se estuviera llevando a cabo tal reforma, esto es, reestablecer un diseño experimental e informes cuantitativos de los juicios cualitativos. (Para una discusión más extensa de los métodos cualitativos-cuantitativos, ver Campbell, 1974).

EL ESTUDIO DE CASOS

El modo de estudio dominante en antropología, ciencia política comparada y sociología comparada continúa siendo el estudio intensivo de un sólo marco particular extraño por una persona que no pertenece al grupo, para quien esta es la única forma intensiva de experiencia de la cultura extranjera. Tales estudios pueden ser escritos por científicos sociales entrenados o por observadores «amateurs» (como misioneros, diplomáticos, reporteros, hombres de negocios, soldados o turistas), cuyas observaciones y ocio pueden incitarles a escribir sobre lo extranjero a su propia cultura. Incluso cuando estos observadores amateurs no escriben, participan fuertemente en el aculturamiento de los científicos sociales dentro de la cultura extranjera, o en la expatriada ubicación en aquella cultura (Kidder, 1971). Otro género similar es la descripción del propio país hecha mientras —o después de que— se está residiendo en otro país, como Kenyatta (1938) describiendo el Kikuyu mientras estaba en Inglaterra como estudiante de Malinowski. Tal conocimiento, escrito o no, lo utilizaré para representar el «conocimiento de sentido común» para la ciencia social comparada. Si conseguimos una correlación cuantitativa significativa de 100 naciones, es porque dependemos de esta clase de conocimiento en todo momento, no por reemplazarla por una metodología cuantitativa, «científica» que sustituya a dicho conocimiento. La generalización cuantitativa multinacional contradecirá al caso único y anecdótico, a la observación naturalista, en alguno puntos, pero lo hará desde la confianza en gran parte en tales anécdotas, casos particulares y observaciones naturalistas.

No quiero decir que este tipo de observación naturalista de sentido común sea objetiva, fiable o imparcial. Pero es todo lo que tenemos. Es el único camino para conocer, aunque sea una forma ruidosa, falible y sesgada. Deberíamos ser conscientes de su debilidad, pero debemos confiar en ella si queremos realizar el proceso de la ciencia social comparada (o monocultural). Volveré sobre el sesgo después, pero antes, permítanme corregir algunos de mis propios excesos anteriores al describir la perspectiva del estudio de casos. La caricatura del estudio de casos concreto que tengo en mente consiste en un observador que se da cuenta de una sorprendente característica de una cultura y dispone del resto de las diferencias entre las demás variables para encontrar una explicación. Puede disponer de prácticamente todos los conceptos causales en su terminología entre los que escoger. Es inevitable que encuentre «una explicación que parezca encajar perfectamente», dada su total falta de «grados de libertad» (Es como si tratara de encajar dos puntos de observación con una fórmula que incluyera miles de términos que se ajustarán de cualquier forma, cuando en buena ciencia deberíamos tener menos términos en nuestra fórmula que puntos de observación).

Esta idea fue expresada por Campbell y Stanley (1966: 6, 7) como sigue (es una cita legítima dado que muchos consideran las comparaciones transculturales, como una fórmula débil de diseño cuasi-experimental; e.g. Lijphart, 1971: 683-685; Boesch y Eckensberger, 1969):

Diseños pre-experimentales

1. «El estudio de un solo caso»

Gran parte de la investigación se efectúa hoy conforme a un diseño en el cual un grupo particular es estudiado una sola vez, después de que se presume que algún agente o tratamiento causa un cambio. Tales estudios tienen el siguiente diagrama:

X O

Como se ha señalado (e.g. Boring, 1954; Stouffer, 1949) tales estudios carecen totalmente de control para tener un mínimo valor científico. El diseño es introducido aquí como un punto mínimo de referencia. Se hace necesario comentarlo a la vista de que aún se invierte en tales diseños y aún se derivan de ellos inferencias causales. Es esencial para la evidencia científica (y para todo

proceso de conocimiento-diagnóstico, incluyendo la retina de los ojos) el proceso de comparación, o registro de diferencias o de contrastes. Cualquier apariencia de conocimiento absoluto o conocimiento intrínseco sobre objetos singulares aislados resulta ilusoria después de efectuar los análisis correspondientes. Asegurarse evidencia científica lleva implícito hacer al menos una comparación. Para que tal comparación sea útil, ambos lados de la comparación deberían hacerse con cuidado y precisión similar. En los estudios de casos tipo Diseño 1, implícitamente se compara un solo caso cuidadosamente estudiado con otros sucesos casualmente observados y recordados. Las inferencias están basadas en expectativas generales sobre cómo habrían sido los datos si X no hubiera ocurrido, etc. Tales estudios a menudo llevan consigo recogidas tediosas de detalles específicos, atenta observación, comprobaciones y similares y, en tales ejemplos, domina el error de la «precisión mal orientada». ¡Cuanto más valioso sería el estudio si las observaciones fueran reducidas a la mitad y el esfuerzo ahorrado dirigido hacia el estudio con igual detalle de un ejemplo de comparación apropiado! Parece muy poco ético en la actualidad permitir estudios de casos o tesis doctorales en educación de esta naturaleza (p.e. un grupo particular observado una sola vez). Los tests «standarizados» en estos estudios proporcionan solamente una ayuda muy limitada, ya que las fuentes rivales de diferencia distintas a X son tan numerosas como para hacer al grupo de referencia «standard» casi inútil como «grupo de control». De la misma manera, muchas fuentes de diferencia incontroladas entre un estudio de un caso actual y futuros potenciales que pudieran ser comparados con él, son tan numerosos como para hacer también inútil su justificación en términos de proporcionar una referencia para futuros estudios. En general, es mejor distribuir el esfuerzo descriptivo a ambos lados de una comparación interesante.

Mi fuerte rechazo de los estudios de un sólo caso o de una comparación de dos partes fue también expresada en mi capítulo del libro de Hsu (1961) como sigue:

«La imposibilidad de interpretar comparaciones entre sólo dos casos naturales. Resulta imperdonable que Malinowski no repitiera las observaciones que hizo a la vista de la importancia de la negación por parte de Malinowski de la interpretación amor-odio del conflicto

de Edipo. Sin embargo, a través de su trabajo de campo sobre otros puntos, la evidencia publicada sobre este tema es realmente escasa. Aunque él hace alusión a pruebas basadas en el contenido manifiesto de sueños, lo que necesitamos son muestras sustantivas de registros detallados de los sueños de chicos y chicas, hombres y mujeres.

Aunque hay una perentoria necesidad de verificar y extender la evidencia de Malinowski sobre las actitudes intrafamiliares de los Trobriand, tales repeticiones tienen una importancia relativa para contrastar las hipótesis freudianas. Aquellos que estemos interesados en utilizar estos datos para dibujar procesos en vez de descubrir exhaustivamente casos únicos debemos aceptar esta regla:

Ninguna comparación de un único par de objetos naturales es interpretable. Entre Trobriand y Viena hay muchas diferencias que podrían constituir explicaciones alternativas potenciales y carecemos de medios para desecharlas. En las comparaciones de parejas, la cláusula «ceteris paribus» es insostenible. Pero no es preciso que se termine aquí la recogida de datos. Tanto el sistema avuncular como el europeo están tan ampliamente distribuidos sobre la tierra que si nuestro objetivo fuera comprobar las teorías del Edipo, podríamos seleccionar una docena de parejas de tribus equiparadas de áreas culturales muy distintas, haciendo que cada par fuera igual a la otra en todo menos en qué adulto varón educa y socializa al niño. Suponiendo que los sueños recogidos de los niños mostraran las diferencias esperadas en cada par, cuantos más pares tuviéramos, menos hipótesis alternativas rivales quedarían y tanto más seguros estaríamos de la confirmación de nuestra hipótesis (Campbell, 1961: 344-345).

Una tercera cita:

«Desde este punto de vista, el problema del "laboratorio" de las relaciones internacionales actuales es la existencia de pocos actores, pocas situaciones, pocas actividades recurrentes en las que conjuntos de actores clasificables como comparables empiecen a interactuar desde puntos de arranque comparables y demasiadas consideraciones teóricas relevantes. Los grados de libertad para la predicción a posteriori, para comprobar la teoría en el conjunto de datos que la sugirió, son menores que en el caso de las predicciones (parte de esta característica se encuentra en la distinción entre tests de significación de una cola frente a tests de significación de dos colas). Estos problemas son tan reales y tan difíciles

para los ensayos cualitativos del científico político o historiador como para el estadístico multivariante. Comunicar el problema a este último puede ser más fácil (aunque la mayoría de las veces suficientemente difícil si afirma estar describiendo todo el universo) dado que éste ha hecho explícitas el número de variables mientras que el estudioso dispone de un potencial de «consideraciones» muy grande y no explícitas al que puede recurrir en un caso o en otro. (Este, y no sus características de reconocimiento de pautas, es el centro de los problemas del enfoque de la comprensión —una fuente de su exquisito y satisfactorio ajuste a casos concretos y de su escasa satisfacción como proceso de contrastación de la realidad») (Raser et al., 1970: 186-187).

Aunque es probable que tengan muchos de estos defectos muchos estudios de casos que implican o mantienen una interpretación o explicación, o que relacionan el caso con la teoría, ahora me parece claro que no todos los tienen, ni tienen porqué tenerlos y que me olvidé de una fuente importante de disciplina (es decir, los grados de libertad, empleando este concepto estadístico para problemas análogos en situaciones no estadísticas). En un estudio de casos hecho por un científico social alerta que tenga un conocimiento serio de lo local, la teoría que utiliza para explicar la diferencia objeto de interés genera también predicciones o expectativas en docenas de otros aspectos de la cultura y éste no retiene su teoría a no ser que queden confirmadas la mayoría. En cierto sentido, ha contrastado su teoría con los grados de libertad resultantes de las múltiples implicaciones de cualquier teoría. El proceso es una especie de emparejamiento de pautas en el que existen muchos aspectos de la pauta exigida por la teoría disponibles para ser emparejados con observaciones del contexto local.

Las experiencias de los científicos sociales confirman este proceso. Incluso en un estudio cualitativo de un sólo caso, el científico social responsable muchas veces no encuentra ninguna explicación satisfactoria. Tal resultado sería imposible si fuera cierta la caricatura del estudio de un sólo caso presentada en las tres citas ofrecidas —existirían en cambio un exceso de explicaciones subjetivas—. Aunque no me cabe la menor duda de que existe un sesgo estadísticamente significativo que favorece el extraer conclusiones en vez de posponer el juicio a la vista de pruebas esencialmente aleatorias (Campbell, 1959; comprobado para los animales por Tol-

man y Krechevsky, 1933; y por la investigación de B.F. Skinner sobre la superstición entre las palomas) quizá este no sea un sesgo dominante, puesto que tanto la evolución biológica como la social habrían eliminado este credulidad en favor de mutantes más discriminables. Es decir, nuestro mecanismo de conocimiento de sentido común debe tener un valor de adaptación, al menos dentro de la ecología en que surgió.

Becker (en comunicación personal; ver también Becker, 1970: 39-62, 25-38) asegura que, casi de un modo invariable, el científico social que lleva a cabo el estudio de un caso intensivamente (mediante observación participante u otro enfoque cualitativo de sentido común para obtener conocimientos), termina encontrando que sus teorías y creencias previas eran erróneas. Si es verdad, este es un hecho importante que debería documentarse sistemáticamente. Si es verdad, queda demostrado que el estudio intensivo de un caso intercultural tiene disciplina y capacidad para rechazar teorías que se dejan de lado en mi caricatura de este método.

Naroll (1962), uno de los archi-cuantificadores en antropología, logra un instrumento poderoso para controlar la calidad de los datos en los estudios cuantitativos mediante la clasificación de la «calidad» del etnógrafo. Merece la pena señalar que los criterios de calidad provienen no del uso por el etnógrafo de un instrumento de la ciencia social cuantitativa (como procedimientos de muestreo aleatorio, cuestionarios estandarizados, tests psicológicos, etc.) sino del conocimiento cualitativamente superior de la cultura descrita, por ejemplo, mediante un mayor tiempo de residencia y un mejor conocimiento del idioma local.

Puede servir de ilustración del principio de disciplina y grados de libertad originados por implicaciones teóricas múltiples el relato de algunas observaciones de primera mano en las primeras fases de un famoso estudio de un caso. En el otoño de 1940 en Berkeley, estaba yo estudiando en un seminario sobre cultura y personalidad dirigido por A.L. Kroeber y E.H. Erickson dedicado a los indios Yurok de California del norte, entre los que Erickson había pasado una temporada de estudio. De esos materiales, Erickson acabó escribiendo su clásico «Observaciones sobre los Yurok» (1943). Kroeber había estudiado esta tribu 20 años antes y había empujado a Erickson a que los estudiara. Kroeber llevó a Erickson hasta los Yurok y le presentó a sus ami-

gos que aún vivían desde la última vez que estuvo. Característicamente, Kroeber empezó el seminario mediante una descripción de los Yurok, que incluía no sólo su situación, geografía, economía, área de cultura, organización social, artefactos físicos, etc. sino también la afirmación de que eran la corporeización clásica de la estructura freudiana del carácter anal. Me acuerdo cómo describía esta estructura y a los Yurok, con una larga lista de adjetivos que empezaban por «p», probablemente algo así como «parsimoniosos, pedantes, petulantes, paranoicos, ...» aunque me he olvidado de los detalles. En cualquier caso, Kroeber preparó la escena para un ejercicio ortodoxo y confirmatorio de antropología psicoanalítica. Kroeber había sido psicoanalizado y había practicado el psicoanálisis durante varios años, entre 1918 y 1922, en San Francisco (T. Kroeber, 1970). Sus lealtades hacia lo antropológicamente fáctico eran mayores que sus lealtades a la teoría freudiana tal y como se ve en sus famosas revisiones de *Totem y Tabú* (Kroeber, 1920, 1939), pero en esta ocasión su orientación era ortodoxamente freudiana. Es indudable que había empujado a Erickson a visitar a los Yurok en la creencia de que eran el tipo ideal dentro de una tipología caracteriológica freudiana.

Erickson era en aquel entonces rabiosamente ortodoxo freudiano. Había hecho una etnografía psicoanalítica de los Sioux (1939), cuya lactancia prolongada usando, el pecho materno, junto con la prohibición concomitante de morder el pecho de la madre, llevaba a una fijación oral-dental de compensación, mostrada en actos de morder o hábitos nerviosos de morder palos, pajas y uñas. Había ampliado de un modo creativo la teoría freudiana del desarrollo del niño con una tipología elaborada de modos y zonas de fijación libidinal durante la infancia (1950: 44-92) pero esta era una elaboración ortodoxa aunque novel. Si el estudio cualitativo de un caso en un solo contexto cultural estuviera tan carente de grados de libertad como mi caricatura decía, en este caso es seguro que el síndrome anal ortodoxo y freudiano habría sido confirmado. Pero no lo fue. Erickson no pudo encontrar nada en el aprendizaje presente o pasado de la higiene personal de los niños que pudiera producir una fijación anal de ningún tipo. Este es un caso dramático de un estudio de un caso que desconfirma teorías, en los términos del análisis del Lijphart (1971), pero con mucha más capacidad de refutación de lo que este autor permite, puesto que

desprecia la disciplina que proviene de la riqueza de detalles relevantes.

Erickson no sólo abandonó, al menos para los Yurok, la hipótesis del carácter anal, sino que generó otra hipótesis y la probó con los datos sobre los Yurok. De las innumerables soluciones alternativas que se le ocurrieron para resolver los interrogantes de los datos, la mayoría eran insostenibles por una razón o por otra. Hago hincapié en esta frase («por una razón u otra») porque parece ejemplarizar el uso de las implicaciones múltiples, de las observaciones múltiples en una población conceptual, la población de las múltiples implicaciones de toda teoría específica: grados de libertad de múltiples implicaciones. La teoría que retuvo al final fue corroborada por observaciones múltiples y diversas del tipo de las que habían refutado sus otras tentativas de solución. No tengo sitio aquí para exponer de una manera convincente su solución y su implausibilidad inicial se acentúa en la siguiente simplificación: en vez de una fijación oral o anal, los Yurok estaban fijados en el canal alimentario globalmente considerado. Esta tesis organizó muchos de los rasgos desacostumbrados de la cultura, por ejemplo sus preocupaciones por el río Klamath que entraba y salía de su pequeño valle, sus casas con entradas y salidas separadas, sus cuentos para niños, incluyendo el del ruiseñor que se metió en la boca del oso y salió por el ano, etc. La lectura de la monografía original resulta mucho más convincente. (Mi propio trabajo para el seminario, hecho junto con F.M. Geier, fue un esfuerzo para dar sentido psicológico a una colección de 30 ó 40 mitos de los Yurok mediante test proyectivos). La nueva teoría de Erickson, que yo recuerde, no había sido formulada aún cuando se desarrolló dicho seminario. El esfuerzo fue tan frustrante que «regresamos» al humor de los «no graduados» y pusimos un «leit motiv» en la tapa atribuido a Confucio: «Se habla de interpretación de mitos sólo si se habla con las eses».

Este aspecto disciplinante de los estudios de un caso que sean sensible, críticos y conscientes, y que he intentado relacionar aquí con el concepto estadístico de tener suficientes grados de libertad para comprobar el ajuste de una hipótesis, es sin duda uno de los aspectos de los principios de emparejamiento de pautas y dependencia del contexto (Campbell, 1966). Además es probable que esté escribiendo, aunque tarde, el mismo argumento que han formulado Lasswell

(1968) y otros (e.g. Raser, 1969) al insistir en un enfoque configuracional. Queda por hacer un análisis epistemológico y estadístico a fondo de la situación. Este desdecirme parcial, no es sino el principio y con vistas a reabrir el problema es probable que esté presentando aquí una defensa exagerada. Pues es cierto que la mente teórica es capaz de racionalización post-hoc con un resultado cualquiera de una forma increíblemente flexible, y en tales racionalizaciones la interpretación en exceso, la capitalización en el azar y el agotamiento de los grados de libertad ocurren a menudo. Un ejemplo en el área de Erickson se encuentra en el contraste entre la presentación de Kardiner (1939) del material sobre Alor de Du Bois y su propia presentación (1944). Para fortalecer su tema de la privación oral, Kardiner presta gran atención a un niño a quien se le ha dado un globo con orejas para jugar y que trata de chupar una de las orejas siendo fotografiado al hacerlo. Desprecia contraindicaciones más relevantes que incluyen la ausencia de «dedos en la boca». En general, si un estudio cuantitativo cuidadoso no puede confirmar una creencia ampliamente difundida, se debe confiar más en este estudio. Pero éste, a su vez, ha rechazado la creencia de sentido común aunque depende de otras muchas creencias. El conocimiento social normal, evidentemente, está sujeto a muchos sesgos específicos de algunos de los cuales me ocuparé a continuación. Pero es el único camino. Aun cuando lo mejoremos, debemos ir a través de él y construir sobre él.

El testimonio de Becker y el ejemplo de Erickson van en contra de un sesgo bien conocido y generalizado a favor de encontrar aquello que una espera encontrar. Francis Bacon lo puso al descubierto como el primero de los «Ídolos (falsas imágenes) de la tribu». «El entendimiento humano, dada su naturaleza especial, fácilmente presupone un mayor grado de orden e igualdad en los casos que el que realmente encuentra. Cuando formula una proposición, el entendimiento humano fuerza todas las demás cosas para que la den apoyo y confirmación. Es error especial y específico del entendimiento humano estar más motivado y movido por afirmaciones que por negaciones» (Bacon, 1853: la cita es de los *Aforismos*). En lo que respecta a juicios perceptuales de fotos desenfocadas, yo he participado en una clara confirmación de esta tesis (Wyatt y Campbell, 1951) y existen muchas más en la literatura (revisada por Campbell, 1959). Pero, aun-

que sea universal, todavía es un sesgo marginal. En nuestro experimento, el sesgo reducía en un 15 % la exactitud que hubiera habido si no hubieran existido expectativas y desaparecía al entrar las fotos dentro de un foco claro (Wyatt y Campbell, 1951). Este es, evidentemente, un sesgo que comparten la percepción física y la social. A menudo entra en conflicto con un sesgo relacionado con los niveles fluctuantes de adaptación (Helson, 1964) que hacen que un observador se dé cuenta de contrastes nuevos a expensas de las similitudes (Campbell, 1961: 341-344; 1959).

Dado que los estudios de una sola cultura siguen siendo el ejemplo más común en la ciencia social comparada, puede ser bueno ofrecer unas cuantas sugerencias para mejorar la disciplina que tales estudios ofrecen como prueba de su teoría. Estas sugerencias se basan en analogías con los estudios cuantitativos. Tal y como aparecen los estudios cuantitativos en las publicaciones, no son inmunes a estos problemas, como veremos. Pero los estudios cuantitativos implican un esfuerzo por fijar niveles de significación que, a su vez, llevan a tomar conciencia de los aspectos sutiles del problema. En lo que sigue, la recomendación general básica es que el investigador que lleve a cabo su estudio de un solo caso guarde registros explícitos de los aspectos analógicos de sus actividades de solución de problemas.

El problema más conocido en lo que respecta a los grados de libertad en los tests de significación es el número de observaciones con las que se contrasta la hipótesis. Para repeticiones similares dentro de un mismo contexto cultural, el antropólogo está normalmente sobrealerta, por ej. el número de ciudades observadas en el que se da la pauta de liderazgo y aquellas en las que no se da. Lo que se sugiere aquí es que se lleve un registro análogo para las implicaciones de la teoría, teniendo como meta registrar todos los pensamientos de contrastación teórica e investigaciones que adoptan este formato: Si la teoría A es verdad, entonces deben seguirse B, C y D. Si estas implicaciones han motivado a la búsqueda activa, el registro será fácil. Si el test se ha realizado a través de un proceso completamente mental, se perderá gran parte de la implicación. Pero incluso para estas pruebas mentales, una atención autoconsciente al problema puede llevar a un registro mucho más completo que el que las monografías sobre estudios de caso ofre-

cen hoy. La mayor parte de esta actividad interrogadora es sin duda literalmente inconsciente y/o transcurre durante el sueño, tal como arguye Poincaré (1913). No se logrará por tanto un registro completo. Probablemente la omisión más común será el no considerar las implicaciones que no ajustan a la teoría especialmente entre los observadores más orientados a ella. La literatura acumulada de otros estudios y de la crítica proporciona un freno lento y parcial a este estado de cosas. Necesitamos acostumbrarnos a impulsar deliberadamente un proceso contrario consistente en animar a otros expertos a que busquen otras implicaciones teóricas y otros hechos capaces de contradecir o confirmar la teoría. En parte esto ya se hace, pero no de manera que se pueda construir una tabla de «blancos» y «tocados». Otra forma de contrarrestar el inevitable etnocentrismo residual de los científicos sociales podría consistir en invitar a otros científicos sociales de la cultura local a que hicieran «pies de página», con comentarios favorables o contrarios a las tesis que se formulan en la monografía regional. Esto contribuiría también a aumentar la tabla de puntuaciones de las tesis «salvadas» y «tocadas». Por casi cada estudio de caso relevante para la teoría clásica existen predicciones no utilizadas hasta la fecha que podrían servir para la validación cruzada (quizás los mitos de los Yurok sobre los que trabajamos Geier y yo podrían proporcionar un caso para la hipótesis de Erickson). Si es un estudio importante, merece un estudio de casos confirmatorio.

Un segundo aspecto de los grados de libertad de los test de significación que merece la pena tomar prestado es la distinción entre tests de una sola cola y tests de dos colas. En el contexto del estudio de casos, esta idea se traduce en la distinción entre el valor confirmatorio del acuerdo entre implicaciones y hechos cuando la teoría se ha formulado a la luz de los hechos y cuando se ha formulado sin una confirmación previa (en cuyo caso, el valor confirmatorio es mayor). Es evidente que en una tabla de puntuaciones estos dos tipos deberían aparecer por separado. Además, cuando la teoría predice correctamente un hecho que sería muy inesperado desde el punto de vista del sentido común o de otras teorías, u otras culturas, por ejemplo, la confirmación es mucho más convincente que si la predicción es trivial. Este punto no es fácilmente aceptado en la tradición de los tests de significación (Meehl, 1967), pero se encuentra en la historia de la cien-

cia, por ejemplo en la utilización de las fases de la luna de Júpiter por parte de Galileo. Una orientación Bayesiana podría probablemente utilizarse para centrarse en la comparación de probabilidades a priori dada la teoría A, en contraposición con las probabilidades a priori dadas otras teorías. Esto no es lo típico de las aplicaciones de Bayes que normalmente tratan, para cada teoría, las probabilidades a priori y las probabilidades después de la recogida de datos. Sea como fuere se puede ciertamente recomendar que la atención de un epistemólogo/estadístico bayesiano se dirija al estudio de casos de confirmación de teorías. Este problema debería elaborarse separadamente para cada uno de los 6 tipos de estudios de caso de Lijphart (1971) (generar hipótesis, confirmar teoría, desconfirmar teoría, etc.) con especial atención a la diferencia en grados de libertad cuando se ha elegido, por ejemplo, el lugar esperando que la teoría sea confirmada después de una búsqueda explícita o implícita por muchas culturas de un ejemplo tal, frente a lanzarse a ojos cerrados y tropezar con una confirmación. El conocimiento previo de la extremosidad de una o dos variables de relevancia, o de ambas, o de ninguna, todo ello influye sobre el problema de los grados de libertad, es decir, sobre la capitalización del azar. Aquí el análisis de Lijphart resulta muy relevante aunque, como en mis análisis anteriores, se olvide de incluir la contrastación de múltiples implicaciones en un solo caso.

Hay un último problema que plantean los grados de libertad que no se trata todavía adecuadamente en los libros de estadística publicados pero cada vez interesa más. Es el problema de llevar a cabo tests de hipótesis múltiples y luego escribir las conclusiones sobre las hipótesis que sean «estadísticamente significativas» utilizando un test que presuponga que el análisis se hizo con una sola hipótesis. Supongamos que nuestros datos hubieran sido generados solamente mediante un proceso aleatorio. Si estudiaríamos todas las interrelaciones entre 15 variables analizando 105 relaciones bivariadas hipotéticas una de cada 20, o sea, más de 5 serían estadísticamente significativas al 5%. El problema aparece en la literatura como «la tasa de error experimental» (Ryan, 1960), el «problema de las comparaciones múltiples» (Scheffé, 1953) o «análisis exploratorio de datos» (Selvin y Stuart, 1966). En combinación con el problema del número de observaciones, puede verse en

forma especialmente extrema en estudios de ciencia política que analizan la relación entre dimensiones de las naciones: si se dispone de tantas variables como naciones, el coeficiente de correlación múltiple que relacione cualquier variable con el resto será 1, incluso si todos los datos de que se disponga son aleatorios. (Puede deducirse este hecho de la fórmula «reducida» de R; Lord y Novick, 1968: 286). Las R calculadas con muestras pequeñas son sistemáticamente más pequeñas cuando se calculan en nuevas muestras. Esta fórmula estima la reducción en ausencia de validación cruzada. Al utilizar el análisis factorial, se sabe que teniendo una muestra pequeña de naciones se suelen extraer más factores y mayores saturaciones factoriales, debido a la mayor variabilidad muestral del coeficiente de correlación calculado en pequeños N. Esta es la explicación del hecho recurrente de que las estructuras factoriales son más inteligibles y explican más cuando los datos son analizados por continentes, en vez de analizarlos para todo el mundo.

Para hacer frente a este problema, en el estudio de casos que intente general hipótesis, por ejemplo en la segunda parte del estudio de Erickson entre los Yurok, se debe llevar un registro de todas las teorías tomadas en consideración en el proceso creativo de solución de problemas. Para representar los grados de libertad de las implicaciones múltiples, se debe también llevar un registro de las implicaciones que han servido para realizar los tests y la tabla de puntuaciones de «tocados» y «blancos». Estoy convencido de que la teoría de Erickson sobre el canal alimentario es mejor que la correlación múltiple pseudo-perfecta que surge de agotar los grados de libertad disponibles al realizar demasiadas pruebas de hipótesis en demasiados pocos casos o implicaciones. También estoy convencido de que se pueden mejorar los estudios de casos en su aspecto comprobación de teorías y en esta dirección. Si también se reconoce la realidad de la interacción de más de un nivel entre variables (por ej. la relación entre las variables A y B es diferente para los distintos niveles de C.D.E.), se debe aceptar que la plétora de hipótesis pausibles y la escasez o costo de los ejemplos imponen límites reales a las ciencias sociales comparadas. Pero el darse cuenta de ejemplos de implicación en el espacio, que sirve de suplemento a los casos de personas, ciudades, nacio-

nes, eras, etc. hace menor el problema, o al menos un poco menor.

Ofrezco una sugerencia más para mejorar el estudio de casos. El estudio de un solo caso, tal como ha sido analizado, es en realidad una comparación de dos casos: la cultura propia y la foránea. Pero esta comparación es muy asimétrica. Asimétrica en un número de parámetros importante: una de las culturas se aprende de niño sin otra alternativa y se aprende a la vez que se adquieren las presuposiciones y lenguaje locales del conocimiento social. La otra cultura se aprende siendo un adulto extranjero. Los detalles de la nueva cultura objeto de estudio son específicos y focales. Se deja sin explicitar la base de comparación de la propia cultura. Si se hicieran explícitos y focales los rasgos de la propia cultura, si se hubieran estudiado directamente, la comparación explícita sería frecuentemente negada. Pero aunque esto sería una mejora, los rasgos de la propia cultura parecerían todavía más razonables, sensibles, morales e intuitivamente comprensibles. Tenderíamos a ver los rasgos de la cultura extranjera como arbitrarios, extraños, enigmáticos y quizá inmorales. Para evitar esto, hice la siguiente sugerencia en un artículo publicado anteriormente:

Triangulación utilizando el propio sesgo cultural de los observadores. El logro de constructos útiles y «realistas» en la ciencia exige métodos múltiples dirigidos al diagnóstico del mismo constructo desde puntos independientes de observación mediante una especie de triangulación. Esta exigencia es necesaria porque hoy en día se acepta que los datos de los sentidos o las lecturas de registros son el resultado de una transacción en la que tanto el observador (registro) como el objeto de las investigaciones contribuye a moldear la forma de los datos. Si tenemos sólo una única observación, resulta imposible separar el componente objetivo del subjetivo. Sin embargo, si se pueden comparar observaciones realizadas por medio de instrumentos distintos y puntos separados que reflejen «los mismos» objetivos, será posible separar los componentes de los datos debidos al observador (instrumento) de los de lo observado. Resulta que este proceso de desenmarañamiento exige tanto observadores múltiples (métodos) como objetos de estudio múltiples y distintos.

Al aplicarlo al estudio de la filosofía de una cultura, esto implica que nuestro estudio típico de un solo observador/una sola cultura es inherentemente ambiguo. Para cualquier ras-

go del informe resulta equívoco saber si es un rasgo del observador o es un rasgo del objeto observado. Para corregir este estado de cosas el paradigma ideal pudiera el que se ve en la Figura 2(a).

equívoco de especificidades de cada ciudad dentro de dicha cultura, a especificidades de personalidad del antropólogo de dicha cultura, a especificidades de personalidad del antropólogo con la cultura estudiada. (Si sólo se

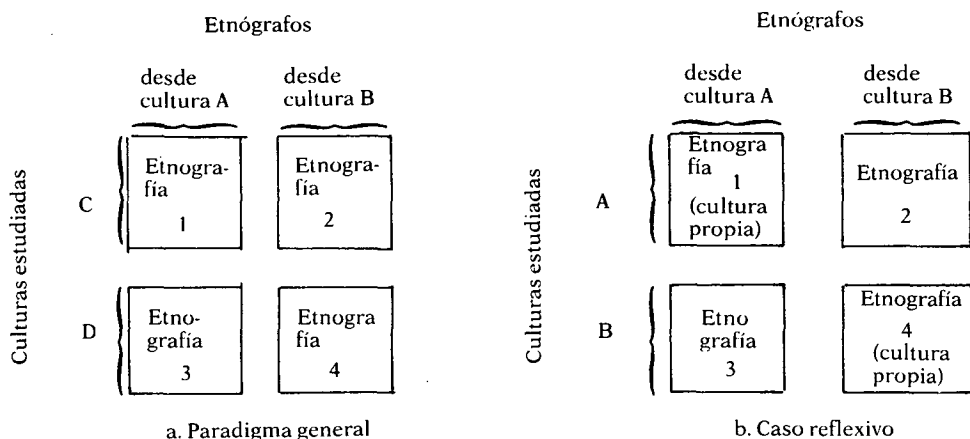


Figura 2: Esquemas de etnografía múltiple para separar el contenido a que contribuye el etnógrafo del contenido de la cultura estudiada.

En el modelo más general, dos antropólogos de distintas culturas estudiarían cada uno una tercera y cuarta culturas. De las cuatro descripciones etnográficas resultantes los atributos comunes en las etnografías 1 y 3 no compartidos por la 2 y 4 podrían ser atribuidas al antropólogo A, los atributos comunes a 2 y 4 que no estuvieran presentes en ninguna otra parte al antropólogo B. Analizando las consistencias entre filas, los tributos comunes a las etnografías 1 y 2 no presentes en las 3 y 4 podrían atribuirse a la cultura C como algo objetivo. Unos atributos comunes a las cuatro etnografías son en sí mismo ambiguos, y se pueden interpretar bien como sesgos compartidos por los etnógrafos o cultura compartida por las culturas estudiadas. Debe señalarse a este respecto que es de desear que se comparen antropólogos de antecedentes culturales lo más dispares posible. En la medida en que los antropólogos vengan de la misma cultura, la repetición de los resultados se convierte más en un problema de fiabilidad que de validez, tal y como se utilizan estos conceptos en los tratados de psicometría. Si el estudio se llevara a cabo utilizando cuatro antropólogos, dos de cada cultura A y B, que estudiaran ciudades separadas de las culturas C y D para evitar interferencias y choques, los atributos característicos exclusivamente de una sola de las etnografías sería atribuibles a un conjunto

utilizara un antropólogo de cada cultura y si cada uno estudiara, haciendo turnos el mismo pueblo de las culturas objeto de estudio, las rasgos específicos de cualquiera de las cuatro etnografías serían debidos equivocadamente a las interacciones cultura-antropólogo, efectos del orden temporal en que el antropólogo reacciona de modo distinto a su segunda cultura, efectos de orden temporal en los que la sociedad reacciona diferentemente ante su segundo estudioso, tendencias históricas e interacciones entre estos fenómenos). La presencia de estas indeterminaciones no debería suprimir ni debería permitirse que hiciera sombra a las grandes ganancias en inteligibilidad originadas por la utilización de múltiples antropólogos.

Aunque la multiplicidad tanto de culturas de antropólogos como de culturas estudiadas es un ideal, también se ganaría si tan solo se lograra la mitad de la parte superior de la Figura 2 (a), es decir, dos culturas de antropólogos centradas en el estudio de una sola cultura a estudiar. En tales triangulaciones, nos enfrentamos de nuevo a la paradoja de la incapacidad de utilizar las diferencias cuando sean tan dominantes que resulte imposible emparejar los aspectos correspondientes del informe objeto de la comparación. La necesidad de este denominador común proporciona justificación a la petición de Hockett de incluir deta-

lles culturales, materiales y conductuales incluso en etnografías que se centren en la determinación de la filosofía de las culturas.

Otra versión del diseño multi-antropólogo multi-objeto de estudio es aquel en el que dos culturas se estudian mutuamente, tal y como aparece en el diagrama de la Figura 2(b). Normalmente el centro de interés son las etnografías 2 y 3, el informe de A sobre B y el de B sobre A. No obstante, de un modo implícito, la descripción de A sobre A y la de B sobre B se pone como base de referencia. Probablemente existe algo de ganancia científica en estos informes, incluso al nivel de conjuntos mutuos de estereotipos. Una vez que se quita el componente evaluativo (cada tribu se ve a sí misma como la mejor), estos conjuntos de estereotipos mutuos muestran un acuerdo digno de mención en la confirmación de la dirección de las diferencias entre grupos» (Campbell, 1964: 331-333).

En línea con la actual discusión, quiero ampliar los requisitos pidiendo que, en una segunda fase del trabajo de campo, se pida a cada etnógrafo que haga validación e invalidación cru-

zada de la interpretación dada por el otro de la cultura que han estudiado ambos.

COMENTARIO FINAL

Este artículo es, obviamente, exploratorio. Es el extremo del péndulo de mi menosprecio anterior y dogmático de los estudios de casos a otro extremo igualmente unilateral. Aunque no va a afectar de un modo apreciable a mi propia enseñanza sobre diseños cuasi-experimentales y métodos de investigación, por el momento suena bien. Después de todo, el hombre es, en su actuación normal, un conocedor muy competente y el conocimiento cualitativo de sentido común no es sustituido por el conocimiento cuantitativo. Más bien, el conocimiento cuantitativo tiene que confiar y apoyarse en el conocimiento cualitativo, incluyendo la percepción ordinaria. (Campbell, 1974). Nosotros, los metodólogos, debemos lograr una epistemología aplicada que integre a ambos.

Referencias

- BACON, F. (1853) «Novum organum», en J. Devey (tans.) *The Physical and Metaphysical Works of Lord Bacon* Londres. H.G. Bohn (publicados originalmente en 1620)
- BECKER, H. (1970) *Sociological Work*. Chicago: Aldine.
- BOESCH, E. E. and L. H. Eckernsberger (1969) «Metgodische Probleme des interkulturellen Vergleichs», pp. 515-566 en C. F. Graumann (ed.) *Socialpsychologie*. Gotingen: Verlag für Psychologie.
- BORING, E. G. (1954) «The nature and history of experimental control». *American Journal of Psychology* 67: 573-589.
- CAMPBELL, D. T. (1974) Qualitative Knowing in action research. Kurt Lewin Award Address, Society for the Psychokogical Study of Social Issues, meeting with the American Psychological Association, New Orleans, 1 de septiembre.
- (1969) A phenomenology of the other one: corrigible, hypothetical and critical pp. 41-69 en T. Mischel (ed.) *Human Action: Conceptual and Empirical Issues*. Nueva York: Academic Press.
- (1966) Pattern matching as an assential in distal knowing. K. R. Hammond (ed.) *The psychology of Egon zbrunswik* Nueva York: Holt, Rinehart & Winston, pp. 81-106.
- (1964) Distinguishing differences of perception from failures of communication in cross-cultural studies, F. S. C. Northrop y H. H. Livingston (eds.) *Cross-Cultural Understanding: Epistemology in Anthropopogy*. Nueva York: Harper & Row, pp.308-336
- (1961) The mutual methodological relevance of anthropology and psychology, en F. L. K. Hsu (ed.) *Psychological Antropology: Approaches Culture and Personality*. Homewood, IL: Dorsey, pp. 33-352
- (1959) Systematic error on the part of human links in communications systems. *Information and Control* 1: 334-369.
- y J. C. STANLEY (1966) *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- DUBOIS, C. A. (1944) *The People of Alor* Minneapolis: University of Minnesota Press.
- ERIKSON, E. H. (1950) *Childhood and Society* Nueva York: W. W. Norton.
- (1943) *Observations on the Yurok*. Publications in American Archaeology and Ethnology, 35, No. 10. Berkeley: University of California Press.

- (1939) Observations on Sioux education. *Journal of Psychology* 7: 101-156.
- GUTTENTAG, M. (1973) Subjectivity and its use in evaluation research. *Evaluation* 1: 60-65.
- (1971) Models and methods in evaluation research. *Journal the Theory of Social Behavior* 1: 75-95.
- HELSON, H. (1964) *Adaptation-Level Theory*. Nueva York: Harper & Row.
- HSU, F.L.K. (ed.) (1961) *Psychological Anthropology: Approaches to Culture and Personality*. Homewood, IL: Dorsey.
- KARDINER, A. (1939) *The individual and His Society*. Nueva York: Columbia University Press.
- KENYATTA, J. (1938) *Facing Mount Kenya*. Londres: Secker & Warburg.
- KIDDER, L. H. (1971) Foreign visitors: a study of the changes in selves, skills, and attitudes of Westerners in India. Tesis doctoral. Northwestern University.
- KROEBER, A. L. (1939) Totem and taboo in retrospect. *American Journal of Sociology* 45: 446-451.
- (1920) Totem and taboo: an ethnological psychoanalysis. *American Anthropologist* 22: 48-55.
- KROEBER, T. (1970) *Alfred Kroeber: A Personal Configuration*. Berkeley: University of California Press.
- LASSWELL, H. D. (1968) The future of the comparative method. *Comparative Politics* 1: 3-18.
- LIJPHART, A. (1971) Comparative politics and the comparative method. *American Political Science Review* 65: 682-693.
- LORD, F. M. y M. R. NOVICK (1968) *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- MEEHL, P. E. (1967) Theory-testing in psychology and physics: a methodological paradox. *Philosophy of Science* 34: 103-115.
- NAROLL, R. (1962) *Data Quality Control: A New Reseach Technique*. Nueva York: Free Press.
- POINCARÉ, H. (1913) Mathematical creation. En H. Poincaré, *The Foundations of Science*. Nueva York: Science Press.
- RASER, J. R. (1969) *Simulation and Society*. Boston: Allyn & Bacon.
- , D. T. CAMPBELL y R. W. CHADWICK (1970) Gaming and simulation for developing theory relevant to international relations. *General Systems*. 15: 183-204.
- RYAN, T. A. (1960) Significance tests for multiple comparisons of proportions, variances and other statistics. *Psychological Bulletin*. 57: 318-328.
- SALASIN, S. (1973) Experimentation revisited: A conversation with Donald T. Campbell. *Evaluation* 1: 7-13.
- SCHEFFE, H. (1953) A method for judging all contrasts in analysis of variance. *Biometrika* 40: 87-104.
- SEGALL, M. H., D. T. CAMPBELL, y M. J. HERSKOVITS (1966) *The influence of Culture on Visual Perception*. Indianapolis: Bobbs-Merrill.
- SELVIN, H. C. y A. STUART (1966) Data-dredging procedures in survey analysis. *American Statistician* 20: 20-23.
- STOUFFER, S. A. (ed.) (1949) *The American Soldier*. Princeton: Princeton University Press.
- TOLMAN, E. C. y I. KRECHEVSKY (1933) Means-end-readiness and hypothesis. *Psychological Review* 40: 60-70.
- WEISS, R. S. y M. REIN (1970) The evaluation of broad-aim programs: experimental design, its difficulties and an alternative. *Administrative Science Quarterly* 15: 97-109.
- (1969) The evaluation of broad-aim programs: a cautionary case and a moral. *Annals of the American Academy of Political and Social Science* 385: 133-142.
- WYATT, D. F. y D. T. CAMPBELL (1951) On the liability of stereotype or hypothesis. *Journal of Abnormal and Social Psychology* 46: 496-500.