

**Mejora de la calidad de la gestión de datos clínicos a
través de sistemas computerizados.**
Aplicación a la informatización de la entrevista
diagnóstica estructurada DICA-IV

Rosario GRANERO PÉREZ

Tesis doctoral dirigida por los Doctores:

Josep M. Doménech i Massons
Lourdes Ezpeleta Ascaso

Departament de Psicobiologia i Metodologia de les Ciències de la Salut
Facultat de Psicologia
Universitat Autònoma de Barcelona

Este trabajo ha sido posible gracias a las ayudas DGICYT PM91-0209 y PM95-126 del
Ministerio de Educación y Cultura

Bellaterra, 1998

Presentación	5
1. Introducción: modelos e instrumentos básicos de clasificación diagnóstica	12
1.1. Presentación del capítulo.....	12
1.2. Las dos grandes orientaciones diagnósticas que han guiado el proceso de clasificación en psicopatología	13
1.2.1. El diagnóstico como sistema de clasificación categorial.....	14
1.2.2. El diagnóstico dimensional.....	20
1.3. Una visión de conjunto.....	23
2. El ordenador como herramienta de ayuda para la evaluación psicológica y la toma de decisiones clínicas	24
2.1. Introducción	24
2.2. La polémica clínico-actuarial: una confrontación que ilustra la necesidad de adoptar herramientas de ayuda en el proceso de psicodiagnóstico	24
2.2.1. Principales acercamientos teóricos al estudio del juicio clínico surgidos tras la confrontación clínico-actuarial.....	27
2.2.2. La adopción de fórmulas: una aplicación de los modelos normativos para simular el juicio clínico en psicodiagnóstico	32
2.3. Introducción de programas informáticos convencionales como herramientas de ayuda en la evaluación psicológica	37
2.3.1. Fases de introducción del ordenador como herramienta de ayuda en el ámbito de la evaluación psicológica	38
2.3.2. Estudio de la equivalencia entre las pruebas tradicionales y su adaptación informatizada	46
2.3.3. Tests adaptativos informatizados en evaluación psicológica	52
2.3.4. El ordenador como herramienta de ayuda para simular el proceso de diagnóstico: los sistemas expertos	53
2.3.5. Diagnóstico clínico a través de redes neuronales	66
2.4. Conclusiones del capítulo	72
2.4.1. Los clínicos, aunque "expertos", cometen errores cuando gestionan los datos y cuando realizan valoraciones psicodiagnósticas	72
2.4.2. La introducción de herramientas informáticas mejora la gestión de datos clínicos y la calidad de los diagnósticos.....	74

3. El ordenador como herramienta de ayuda para mejorar la calidad de la gestión de datos clínicos	76
3.1. Introducción	76
3.2. Fuentes de error relacionadas con la no observación	80
3.2.1. Fuentes de error debidas al observador/entrevistador	80
3.2.2. Fuentes de error debidas al entrevistado.....	81
3.2.3. Fuentes de error debidas al instrumento de evaluación.....	84
3.2.4. Fuentes de error debidas a la forma de recogida de datos	85
3.3. Fuentes de error relacionadas con la gestión de los datos	87
3.3.1. Base de datos, sistemas de bases de datos y sistemas de gestión de bases de datos..	88
3.3.2. Cómo mejorar la calidad de la gestión de datos desde la óptica del diseño de bases de datos	92
3.3.3. Mejora de la calidad a través de la definición de la estructura de la base de datos...	93
3.3.4. Mejora de la calidad durante la grabación de los datos	94
3.3.5. Mejora de la calidad a través del diseño de la interfaz.....	101
3.3.6. Mejora de la calidad durante la fase de preparación de la matriz de datos para ser analizada: creación y modificación de variables.....	106
3.3.7. Mejora de la calidad durante el análisis estadístico.....	108
3.3.8. Cómo se puede evaluar el proceso de calidad de la gestión de datos clínicos	109
3.3.9. Síntesis de las estrategias de control propuestas	113
4. Un estudio de la calidad de la captura de datos psicodiagnósticos.....	114
4.1. Introducción: justificación de la experiencia y objetivos	114
4.1.1. Primer objetivo de la experiencia	115
4.1.2. Segundo objetivo de la experiencia.....	115
4.2. Método	116
4.2.1. Sujetos	116
4.2.2. Material.....	116
4.2.3. Diseño.....	117
4.2.4. Procedimiento.....	118
4.2.5. Análisis de los datos	119
4.3. Resultados	126
4.3.1. Estudio de las discrepancias entre las Fases A y B1.....	126
4.3.2. Estudio de los errores entre los datos verificados y el patrón.....	135
4.4. Discusión.....	149

5. Una propuesta para mejorar la calidad de la gestión de datos clínicos: aplicación a la informatización de la entrevista estructurada DICA-IV	152
5.1. Introducción: justificación de la experiencia.....	152
5.2. Descripción del protocolo escogido para adaptar en formato informatizado.....	154
5.3. Descripción del software escogido para construir la base de datos e informatizar la entrevista DICA-IV	156
5.4. Diseño de la estructura de la base de datos	157
5.5. Creación de la estructura de entrada de datos	160
5.5.1. Definición de las pantallas de captura de datos	160
5.5.2. Establecimiento de restricciones entre los datos	164
5.5.3. Facilitación de la recogida simultánea de subréplicas.....	170
5.5.4. Definición de los indicadores de salud mental	171
5.6. Valoración del uso de entrevistas estructuradas informatizadas en psicodiagnóstico	183
5.6.1. Ventajas del uso de entrevistas informatizadas	183
5.6.2. Limitaciones que se han relacionado con el uso de entrevistas informatizadas	186
5.7. Propuestas para informatizar entrevistas estructuradas en psicopatología.....	191
5.7.1. Disponibilidad de una entrevista con diferentes formatos según la finalidad de la evaluación	191
5.7.2. Adaptación de la entrevista a diferentes idiomas o a diferentes niveles de comprensión del vocabulario	192
5.7.3. Adaptación de la entrevista en formato autoadministrado	193
5.7.4. Definición de indicadores de salud mental categoriales y dimensionales	194
5.7.5. Obtención de informes finales completos	196
6 Discusión	197
6.1. Los sistemas informáticos ganan protagonismo en evaluación psicológica	197
6.2. Las entrevistas estructuradas informatizadas mejoran la calidad de la gestión de datos	199
6.2.1. Mejora de la evaluación adoptando reactivos adecuados	201
6.2.2. Mejora de la evaluación definiendo una estructura de captura de datos protegida .	202
6.2.3. Mejora de la evaluación adecuando la gestión de los datos	202
6.2.4. Reflexiones finales y futuras líneas de trabajo	204
Referencias.....	207

100
100

PRESENTACIÓN

Cuando se afirma que las enfermedades mentales constituyen un hecho contemporáneo, el consenso acostumbra a ser unánime. La actual vigencia clínica de los trastornos psicológicos deja constancia en múltiples estudios epidemiológicos que revelan que en nuestra sociedad un porcentaje importante de la población presenta dichas alteraciones, y que otro porcentaje también considerable muestra diferentes problemas derivados de constantes como el estrés, la presencia de enfermedades crónicas, la desestructura familiar o una situación económica precaria (Buckner y Bassuk, 1997; Keenan, Shaw, Walsh, Delliquadri y Giovannelli, 1997; Martin, Kupersmidt y Harter, 1996). Los informes epidemiológicos también constatan que los trastornos mentales aparecen a cualquier edad, si bien en su mayoría se relacionan con el momento evolutivo en el que se presentan (Esser, Schmidt y Woerner, 1990; Fischer, Barkley, Fletcher y Smallish, 1993; Grilo et al., 1996; Hankin et al., 1998; Mattanah, Becker, Levy, Edell y McGlashan, 1995; Roberts, Attkisson y Rosenblatt, 1998; Verhulst y Van der Ende, 1992). Así, por ejemplo, en la población general adulta de la mayoría de países desarrollados los valores de prevalencia se sitúan cercanos al 25% (Ferdinand y Verhulst, 1996), y en la población infantil y juvenil parece acertado aceptar índices que oscilan entre el 18% y el 27% (Bird, 1996; Costello, Burns, Angold y Leaf, 1993; McDermott, 1996; McGee et al., 1990; Verhulst, Van der Ende, Ferdinand y Kasius, 1997; Wiemann, Berenson, Wagner y Landwehr, 1996). En España, el informe epidemiológico no resulta más esperanzador. Se estima que entre un 20% y un 35% de los españoles han sufrido algún trastorno psicológico durante los últimos años (Chocrón Bentata et al., 1996; Ruiz, Pérez y Vicente, 1997), y que en un porcentaje alto de estos casos el episodio ha resultado lo suficientemente grave como para requerir consulta y/o hospitalización. Asimismo, y en relación con los datos epidemiológicos disponibles, es importante tener presente algunos estudios que revelan que durante los últimos años se está experimentando un temible aumento en estos trastornos (Achenbach y Howell, 1993), y también que el curso de dichas alteraciones tiende a prolongarse durante bastante tiempo (Biederman et al., 1996; Steinhausen, Meier y Angst, 1998).

Las cifras epidemiológicas señaladas (a modo orientativo) en el párrafo anterior constatan la importancia de estimar con precisión tanto la frecuencia de los trastornos mentales como su génesis y su distribución en las poblaciones, ya que sólo así es posible formular "hipótesis sólidas" sobre los factores de riesgo y establecer bases sobre las cuales diseñar los correspondientes programas de prevención y tratamiento (King, Hovey, Brand y Ghaziuddin, 1997). Pero determinar estos índices de frecuencia y detectar los correspondientes correlatos nunca ha resultado una tarea fácil. Durante años, sólo una parte de los sujetos que presentan este tipo de problemas ha sido identificada correctamente en la evaluación psicodiagnóstica que se lleva a cabo en los servicios de psicología y/o psiquiatría. En la actualidad, todavía muchos especialistas y responsables de la gestión sanitaria se lamentan de la falta de precisión con la que se manejan los datos de frecuencia y riesgo cuando se refieren a los trastornos psicológicos en la población.

Aunque gran parte de las dificultades planteadas en torno al estudio epidemiológico de los problemas psicológicos se debe a la conjunción de múltiples factores, diferentes trabajos coinciden al subrayar cuatro hechos fundamentales que dificultan la correcta identificación de los diagnósticos y de los factores de riesgo (Jellinek y Murphy, 1988). En primer lugar, si tenemos en cuenta que uno de los métodos comúnmente utilizados consiste en realizar los estudios con los sujetos que han acudido a los servicios de salud mental, se comete un importante sesgo al excluir todos aquellos individuos que, aún cuando padecen estos

trastornos, no acuden a las consultas ni se someten a ningún tratamiento (especialmente cuando experimentan sentimientos de vergüenza, miedo, negación o tolerancia). En segundo lugar, la tipificación de las alteraciones biológicas, psicológicas y sociales de cada trastorno es relativa, y está abierta a numerosas excepciones, a la presencia y a la ausencia de síntomas de muy diversa índole. En tercer lugar, la evaluación psicopatológica puede resultar una tarea sumamente compleja, y el resultado obtenido en dicho proceso una mera y cuestionable etiqueta diagnóstica-clasificatoria, sobre todo cuando no se utilizan instrumentos estandarizados adecuados para medir las desviaciones respecto al comportamiento normal o cuando los sujetos encargados de llevar a cabo la evaluación no poseen las suficientes habilidades para realizar estas funciones. Finalmente, los resultados publicados en gran parte de los trabajos aparecidos en el seno de la psicopatología se basan en análisis estadísticos, más o menos complejos, que presuponen la calidad de la información que se analiza. En este último caso, los investigadores pueden introducir sesgos de magnitud desconocida al no contemplar la posibilidad de que a lo largo del proceso que integra una investigación se presenten numerosas oportunidades para introducir errores en los datos (Cobos, 1995).

El trabajo que presentamos se relaciona con los dos últimos puntos expuestos en el párrafo anterior, y se ubica dentro del proceso de evaluación y mejora de los instrumentos, aspectos estrechamente relacionados con el problema de la medición y de la calidad de datos en psicopatología. En concreto abordamos el problema del diagnóstico psicológico y de cómo optimizarlo desde el punto de vista del diseño de pruebas y sistemas de evaluación informatizados. En los siguientes párrafos intentaremos justificar nuestro interés en estos aspectos y por qué consideramos necesario emprender nuevas investigaciones en esta línea de trabajo.

Es evidente que para cualquier clínico el requisito previo al diagnóstico es la evaluación y la medición de las alteraciones. Todo evaluador lleva a cabo una serie de tareas con dos objetivos fundamentales: seleccionar información suficiente que le permita formular los supuestos sobre cada caso concreto y verificar tales hipótesis mediante técnicas de evaluación adecuadas (y, en su caso, por medio de contrastación experimental). Las tareas básicas que permiten alcanzar estos propósitos se pueden concretar en: la recogida (selectiva) de información, la combinación de los datos para otorgarles un sentido psicológico, la toma de decisiones sobre las categorías que resultan más verosímiles a la hora de clasificar las conductas de los sujetos y las predicciones sobre el tratamiento más adecuado y/o la competencia futura de los sujetos (Haynes, 1992). Evidentemente, si el propósito fundamental de nuestros profesionales es que el psicodiagnóstico forme parte del estudio científico del comportamiento humano, la evaluación psicológica debe constituir un procedimiento reglado, con fases establecidas coincidentes con las del método científico-positivo, y concebida de tal forma que pueda ser repetida en condiciones similares por diferentes evaluadores (Fernández Ballesteros, 1994).

Lamentablemente, la evaluación del comportamiento para determinar la presencia de los trastornos siempre ha resultado compleja. Durante muchas décadas, prácticamente todos los aspectos relacionados con el proceso de evaluación eran atribuidos a la responsabilidad de cada psicólogo o psiquiatra particular: escoger las pruebas de evaluación apropiadas en cada caso, determinar la secuencia de administración, combinar la información disponible y emitir juicios o hipótesis sobre los diagnósticos o tratamientos más adecuados. En este contexto, el concepto de medición y uso selectivo de técnicas estandarizadas parecía no jugar un papel importante en el seno de la evaluación psicológica. Es cierto que en determinados aspectos, como por ejemplo la evaluación de la inteligencia o la personalidad, se obtenían estimaciones sobre los supuestos atributos o dimensiones subyacentes. Sin embargo, en el ámbito particular del psicodiagnóstico, la posibilidad de medir de forma objetiva y estandarizada las alteraciones de los sujetos se concibió durante mucho tiempo como un método más, ni más ni menos válido, que el resto de estrategias utilizadas para adquirir información de los sujetos

(Reich, 1992). Una de las herramientas de evaluación más utilizadas en este entorno ha sido, precisamente, la entrevista clínica abierta.

En los últimos años el concepto de medición ha ganado importancia en el ámbito de la evaluación psicodiagnóstica. Este hecho ha impulsado numerosos proyectos cuyo objetivo primordial ha consistido en perfeccionar las técnicas de evaluación de las que disponen clínicos e investigadores, y también en desarrollar nuevos instrumentos de medida. Dentro de este marco de trabajo, recientemente se han comenzado a utilizar las denominadas entrevistas estructuradas y/o semiestructuradas, esto es, protocolos basados directamente en sistemas de clasificación diagnóstica y diseñados especialmente para ofrecer un diagnóstico desde una perspectiva claramente categorial. La característica fundamental de estos instrumentos de evaluación consiste en limitar el número de preguntas, concretar su correspondiente formulación verbal y estandarizar tanto las "rutinas" como los criterios y los algoritmos de corrección. El desarrollo de instrumentos de evaluación estandarizados tales como las entrevistas estructuradas, no sólo ha permitido comparar fácilmente los resultados de los diferentes estudios, sino que ha llevado a muchos profesionales al convencimiento de que en la actualidad existen más métodos diagnósticos capaces de alcanzar una gran precisión (al menos, mucho mayor que la que se conseguía hace pocos años, cuando el clínico aplicaba su propio juicio a los datos obtenidos exclusivamente de una exploración no sistematizada).

Pero para otros muchos profesionales, el desarrollo de instrumentos de evaluación no se acaba con la mera construcción de protocolos estandarizados; bien al contrario, el ingente número de trabajos aparecidos en la literatura indica que el perfeccionamiento de las técnicas de evaluación constituye un proceso en constante desarrollo. Es cierto que, en definitiva, y al igual que otras disciplinas del ámbito de la salud, la psicopatología en general y la evaluación psicodiagnóstica en particular constituyen disciplinas que se orientan cada vez más hacia la instrumentalización y la depuración metodológica. Sin embargo los resultados de numerosas investigaciones indican que todavía queda un gran camino por recorrer. En este sentido, parece que no hemos hecho más que entrar, y sólo a nivel empírico, en una línea caracterizada por la definición de constructos de diagnóstico objetivos y claramente cuantificados. El reto más inminente ahora consiste en llevar estas nuevas propuestas metodológicas a la práctica clínica diaria.

En estrecha relación con los aspectos de estandarización e instrumentalización del proceso de evaluación psicológica, una de las líneas de trabajo más fructíferas en los últimos años consiste en establecer el uso de la informática y de sus prestaciones en el ámbito de la psicopatología y del psicodiagnóstico. Sabemos que hoy en día ya existe una gran cantidad de software de ayuda en el ámbito de la evaluación psicológica. Los diferentes programas de ordenador se pueden clasificar según las fases del proceso de diagnóstico al que pretenden servir de ayuda (Westmeyer y Hageböck, 1992): establecimiento y análisis del problema, proposición de hipótesis, planificación de la recogida de la información, procesamiento de los datos, emisión de un diagnóstico y/o pronóstico, puesta en marcha del tratamiento seleccionado y evaluación de los programas existentes. La mayoría de los programas actuales se han diseñado para ser empleados en las fases de recogida, corrección, análisis e interpretación de las pruebas de evaluación (fundamentalmente escalas, cuestionarios y entrevistas clínicas), y en la actualidad muchos instrumentos tipificados de recogida de información pueden ser aplicados, corregidos e interpretados mediante ordenador (Bloom, 1992; Godoy, 1996; Honaker y Fowler, 1990; Jäger, 1991; Kleinmuntz, 1991; Lewis, 1992).

Las ventajas de disponer de sistemas informatizados, en psicopatología en general y en psicodiagnóstico en particular, resultan evidentes desde cualquier punto de vista. Por ejemplo, si asumimos que el proceso de investigación científica comporta un conjunto de actividades secuenciales que en gran medida gira en torno al manejo o gestión de los datos, es evidente que el uso de sistemas informatizados cuenta con importantes ventajas a la hora de recoger, preparar y analizar datos clínicos y epidemiológicos, además de facilitar la disponibilidad casi

inmediata de un perfil psicopatológico de los sujetos. Desde esta perspectiva, la tecnología informática se considera del todo indispensable cuando se manejan grandes volúmenes de datos. Algunos investigadores han incluso augurado que el avance en disciplinas tales como la evaluación psicológica o el psicodiagnóstico sólo será posible en la medida en que se adopten y utilicen las actuales prestaciones tecnológicas en "software" y "hardware" (Bloom, 1992; Shanks, 1989).

Pero el mero uso del ordenador, al menos en lo que se refiere a la mayoría de los programas que utilizan muchos usuarios en psicopatología, no garantiza en absoluto la adecuada gestión de los datos, ni tampoco que los resultados obtenidos en una investigación reflejen con fidelidad las relaciones que supuestamente representan. Así, por ejemplo, la calidad de la información que se maneja en el ámbito de las Ciencias de la Salud se ha abordado desde teorías como el diseño de bases de datos, esto es, a través de procedimientos derivados, fundamentalmente, del álgebra relacional y de los trabajos de la "Human Computer Interaction" (Carroll, 1993; Dillon, 1983; González, 1993; Johnson, 1992; Pocius, 1991; Schneiderman, 1992; Wallace y Anderson, 1993). Siguiendo esta línea de trabajos, actualmente se conocen diversas fuentes de error que pueden incidir sobre la calidad de los datos, y se asume que muchos de ellos actúan durante las fases de entrada de datos y preparación de la matriz que posteriormente será sometida a análisis estadístico. Por consiguiente, tanto en la fase de diseño de un estudio como la fase de diseño de cada uno de los instrumentos de evaluación se deberían contemplar dos etapas fundamentales: la definición de las potenciales fuentes de error y el análisis de los recursos del ordenador con objeto de erradicarlas. Lamentablemente ambas etapas son comúnmente olvidadas en muchos estudios de investigación que, paralela y sorprendentemente, adoptan la denominación de "científicos".

Los inconvenientes expuestos en el párrafo anterior nos ha llevado a considerar que es preciso recabar en breve nueva investigación para documentar los sesgos que se producen en el proceso de evaluación psicológica cuando se utilizan estas nuevas tecnologías. Asimismo, es fundamental alcanzar a describir y valorar estos efectos sobre la calidad del proceso de evaluación y sobre la validez y utilidad de los procesos de decisión diagnóstica. De este modo, será posible llegar a diseñar, validar e implementar los correspondientes procedimientos de control destinados a incrementar la calidad de los resultados de la evaluación.

Ubicado dentro de esta línea de investigación, este proyecto constituye, ante todo, una propuesta que pretende resultar útil para quienes se enfrentan a la ardua labor de diseñar herramientas de evaluación psicodiagnóstica informatizadas (tanto en contextos clínicos como de investigación). El núcleo fundamental se centra en la definición de las diferentes fuentes de error que pueden estar incidiendo en el proceso de captura de los datos y en la posterior combinación de la información diagnóstica disponible. Paralelamente, se proponen diferentes procedimientos de control que deberían implementarse de forma rutinaria para mejorar, tanto como sea posible, la calidad de los datos de un estudio de investigación.

En los dos primeros capítulos, de carácter teórico, se efectúa un breve recorrido histórico por el proceso de evaluación psicológica y la toma de decisiones diagnósticas, se presentan las herramientas de evaluación más utilizadas en la actualidad (prestando una especial atención al enfoque categorial y al ámbito de aplicación de la población infantil y juvenil), y se delimita cuál ha sido el uso de la tecnología informática en el ámbito de ayuda al proceso de psicodiagnóstico.

En el tercer capítulo, también de carácter teórico, se revisa en qué medida el ordenador ha servido particularmente como herramienta para la mejora de la calidad de los datos clínicos, tomando como punto de referencia la gestión de la información desde el ámbito del diseño de bases de datos relacionales.

En el cuarto capítulo, de carácter empírico, pretendemos estudiar las dificultades y las implicaciones prácticas que tiene la transcripción al ordenador de los datos clínicos que

previamente han sido recogidos (a mano) en formularios de papel. Para ello se ha diseñado un experimento que consiste en grabar en un ordenador, en ausencia de controles automáticos, un total de 380 entrevistas estructuradas que habían sido codificadas en papel. Hemos diseñado esta experiencia porque consideramos que representa la forma habitual que utilizan la mayoría de profesionales en nuestro país a la hora de llevar a cabo el registro y la captura de datos clínicos. El objetivo general del estudio obedece a la medición de los errores de operador, y se descompone en los siguientes objetivos específicos fundamentales:

- Estudio, a nivel cualitativo y cuantitativo, de los errores que se cometen durante el registro de datos que habían sido recogidos en papel. Se presta especial atención al tipo de error (por omisión o cambio de código), a la distribución en función del tipo de preguntas, del tipo de protocolo y del entrevistador, y a las implicaciones en el estatus diagnóstico final de los sujetos.
- Estudio comparativo de dos estrategias diferentes de control de la calidad de los datos orientadas al proceso de verificación: la "verificación aleatoria" frente a la doble entrada (con y sin "feedback").

En el quinto capítulo se presenta una alternativa que ha demostrado mejorar considerablemente la calidad de la gestión de datos clínicos: disponer de instrumentos de evaluación informatizados para la captura directa ("on-line") de la información. En concreto, se presentan las fases que hemos seguido para informatizar una entrevista diagnóstica estructurada, la Diagnostic Interview for Children and Adolescents DICA-IV (Reich, Leacock y Shanfeld, 1997). En el desarrollo del capítulo se presta una especial atención al diseño de la base de datos (de carácter relacional, aplicando un proceso de normalización), al establecimiento de restricciones para la entrada protegida (controles de pre-validación, invalidación y post-validación de datos) y a la implementación de los reactivos y los constructos generados de diagnóstico en formato informatizado. En relación a este último punto, se plantea un problema muy habitual cuando se manejan variables en las que faltan datos de algunos sujetos. Veremos como, por ejemplo, cuando se evalúa un síntoma cuyo campo de registro aparece vacío es fundamental diferenciar entre "valor desconocido", "valor no aplicable", "valor deducible" y "ausencia de síntoma". En este trabajo se proponen dos estrategias fundamentales: proteger la entrada de datos para prevenir la presencia de campos en blanco, ya que de otro modo puede ser imposible definir el "estatus diagnóstico" de los sujetos, y depurar los algoritmos de corrección de las pruebas de evaluación diagnóstica para que el resultado final no confunda las categorías de "diagnóstico ausente" y "diagnóstico no evaluable". En relación a esta última estrategia, puesto que los instrumentos de evaluación habituales en psicodiagnóstico se limitan a la definición de la sintomatología, hemos considerado oportuno expresar en lenguaje formal los diagnósticos DSM-IV a través de macros en sintaxis SPSS y SAS, junto a la lista completa de llamadas, ya que de este modo se estandariza y garantiza la ausencia de errores en el proceso de definición de dichos constructos.

En el quinto capítulo también se valoran las ventajas y limitaciones que conlleva la informatización de entrevistas diagnósticas y la disponibilidad de procedimientos de evaluación psicológica en formato informatizado. La exposición se ilustra con la experiencia obtenida a través de la adaptación de la DICA-IV. Finalmente, este capítulo concluye presentando unas directrices que pueden resultar de interés para quienes diseñan e informatizan instrumentos de evaluación diagnóstica estructurados.

En el sexto capítulo se discuten las implicaciones teóricas y prácticas de este trabajo, y se resumen las conclusiones principales que se derivan de la labor que hemos llevado a cabo.

En síntesis, este trabajo, en conjunto, incide en uno de los aspectos cruciales que se plantean cuando se investiga en Ciencias de la Salud: la dificultad de la medición. En este sentido opinamos que, lamentablemente, el desarrollo científico no ha conseguido todavía alcanzar la meta a la hora de definir los métodos de medición. Pero también creemos que la situación no es tan desalentadora, y que los esfuerzos que se realizan en esta línea ofrecen trabajos de innegable interés clínico y epidemiológico. Deseamos que nuestra investigación aporte también una pequeña contribución a este mar de conocimientos.

Nos gustaría matizar que a lo largo de la exposición utilizaremos frecuentemente conceptos tales como clasificación, taxonomía, diagnóstico, evaluación, juicio, toma de decisiones o elección. Debido a la controversia que suscitan los conceptos implícitos en cada uno de estos términos, creemos conveniente facilitar una breve aclaración sobre el significado que asignaremos a dichos conceptos.

Mediante los términos "clasificación" y "taxonomía" nos referiremos a la definición de las categorías o dimensiones principales de la conducta, en nuestro caso en lo que se refiere a formulación de trastornos psicológicos, tanto en el ámbito de la clínica como en el de la investigación. Con el término "diagnóstico" haremos referencia al proceso (concreto) por el cual se asigna a un individuo a una determinada categoría del sistema de clasificación. El término "evaluación" aludirá al proceso (general) continuo de examinar a los sujetos, en gran medida para ayudar a los procesos de clasificación y diagnóstico. Los términos "juicio clínico" y "razonamiento clínico o diagnóstico" se referirán, indistintamente, al conjunto de tareas o procesos cognitivos que conlleva el proceso de evaluación en general y el establecimiento de un diagnóstico en particular -no distinguiremos si dicho proceso se lleva a cabo de forma global o intuitiva, si se ejecuta de forma explicitable, clara y distinguible, o si se conduce de acuerdo a algún sistema de reglas normativas-. Finalmente, al tratar sobre aspectos más metodológicos utilizaremos el término "valor desconocido" para designar lo que habitualmente se denomina "valor missing".

También nos gustaría aclarar en esta presentación que en este trabajo concebimos la evaluación psicológica como un proceso general de toma de decisiones. Esto supone que, al menos en la delimitación de los aspectos teóricos del trabajo, consideramos de innegable interés conceder una especial atención a diferentes aspectos que se relacionan estrechamente con lo que en la literatura se ha venido denominando "juicio clínico". Sabemos que el problema de la evaluación diagnóstica no se acaba con estandarizar la recogida de la información y la corrección de las correspondientes pruebas. En nuestra disciplina, la evaluación diagnóstica requiere que el profesional interprete la información disponible de los sujetos y le otorgue un significado psicológico. En relación a este punto, en la actualidad los profesionales clínicos disponen de una gran cantidad de conocimientos que pueden consultar y utilizar, por ejemplo las teorías e hipótesis explicativas enunciadas en relación a dichas alteraciones. Lamentablemente, estas bases de conocimientos pueden llegar a ser tan exhaustivas, que muchos profesionales no disponen ni del tiempo ni de los recursos necesarios para poder abarcarlas completamente (Westmeyer y Hageböck, 1992), y en la mayoría de los casos los clínicos se limitan a realizar un uso tan selectivo como restringido de estas fuentes documentales. Asimismo, y en relación al estudio del juicio clínico, también señalaremos que en la actualidad disponemos de una gran cantidad de trabajos que evalúan los diferentes sesgos y errores que pueden afectar al proceso de razonamiento y al juicio de los clínicos (Adelman y Bresnick, 1992; Arkes, 1991; Arocha, Patel y Patel, 1993; Block y Harper, 1991; Christensen-Szalanski y Willham, 1991; Edwards, 1992; Ellis, Robbins, Schult, Ladany y Banker, 1990; Elstein et al., 1990; Fischhoff y Beyth-Marom, 1988; Garb, 1989; Gigerenker, 1991a, 1991b; Gigerenker y Hug, 1992; Gigerenker, Hoffrage y Kleinbölting, 1991; Godoy, 1996; Goldstein, 1990; Hawkins y Hastie, 1990; Harvey, 1992; Johnson, Grazioli, Jamal y Zualkernan, 1992; Joseph y Patel, 1990; Kahneman, 1991; Keren, 1992; Koehler, 1991, 1993, 1994; Miller y Karoni, 1996; Reilly y Doherty, 1992; Paese y Snizek, 1991; Smith y Kida,

1991). En este listado de potenciales fuentes de error se incluyen tanto factores internos al sujeto (por ejemplo, la limitación de la capacidad cognitiva humana y el uso de heurísticos rutinarios) como factores externos (por ejemplo, la complejidad de la tarea). En nuestro trabajo señalamos que, precisamente, una de las formas de responder a las limitaciones de los clínicos (incluso de los que se consideran más expertos) en sus tareas de evaluación y diagnóstico consiste en diseñar, validar e implementar sistemas computerizados que les sirvan de ayuda a la hora de aplicar la base de conocimientos disponibles, tanto para la recogida y la corrección de las pruebas, como para la posterior interpretación de los datos que manejan. Los resultados obtenidos en diferentes trabajos sugieren que esta estrategia suele resultar más adecuada, a nivel de eficacia y eficiencia, que el uso exclusivo de entrenamientos intensivos para los profesionales (Hammond, Hamm, Grassia y Pearson, 1987; Kleinmuntz, 1990).

INTRODUCCIÓN: MODELOS E INSTRUMENTOS BÁSICOS DE CLASIFICACIÓN DIAGNÓSTICA

"Classification systems in psychiatry have been and will continue to be tentative as long as disorders are grouped on the basis of signs and symptoms elicited in interviews. While this is so, it makes sense for researchers to emphasize improved measurement of signs and symptoms that can be combined into diagnoses by logical systems other than those of current DSMs and ICDs. (...) Establishing accurate measures of signs and symptoms will ensure that the measures will not be out of date with changes in rationales for how they should be combined to indicate particular diagnoses".

(Dohrenwend, 1998)

1.1. PRESENTACIÓN DEL CAPÍTULO

En este primer capítulo, de carácter meramente introductorio, exponemos cuáles han sido los modelos teóricos fundamentales que han orientado el proceso de evaluación psicológica, y en particular la etapa de clasificación diagnóstica en psicopatología. La razón de este preámbulo obedece a la idea general de que todo proceso de evaluación clínica se inscribe dentro de un determinado marco teórico de referencia. Este anclaje referencial condiciona dos aspectos fundamentales relacionados con la evaluación: la metodología de obtención de datos para la elaboración del diagnóstico, y las estrategias (y técnicas) que se utilizarán en cada caso.

Nos gustaría destacar que cuando se intenta realizar una revisión de los enfoques fundamentales que han guiado el proceso de evaluación psicodiagnóstica y la construcción de instrumentos de medida nos hallamos ante varios problemas teóricos y metodológicos (Pedreira y Sánchez-Gimeno, 1992; Kraemer et al., 1987).

La primera dificultad alude a la inexistencia de un consenso teórico. A pesar de que en los últimos 30 años se han producido cambios notables en el plano conceptual que reflejan la necesidad de una búsqueda de nuevas perspectivas para entender cómo se produce la adaptación psicológica de los sujetos, parece que todavía no hemos encontrado un modelo teórico lo suficientemente amplio y explicativo de este complejo sistema. Hasta el momento, sólo contamos con abordajes capaces de explicar, con más o menos precisión, algunos aspectos del funcionamiento psicológico. En principio se podría suponer que en los ámbitos de la evaluación y del psicodiagnóstico existen tantas orientaciones teóricas como marcos de trabajo referenciales dentro del seno de la psicopatología (Vallejo, 1998). Sin embargo, también parece existir una cierta uniformidad al considerar que los diferentes anclajes teóricos que han coexistido durante las últimas décadas en esta disciplina se pueden resumir en dos modelos fundamentales: el modelo categorial frente al dimensional. La exposición de este capítulo se centra en el denominado modelo categorial. En primer lugar, porque este modelo ha prevalecido durante mucho tiempo como dominante en el ámbito de la evaluación psicodiagnóstica. En segundo lugar, porque las herramientas de medida con las que nuestro equipo viene trabajando durante los últimos años participan de los supuestos fundamentales de este marco general.

La segunda gran dificultad se relaciona con la expuesta en el párrafo anterior: la pluralidad teórica ha estimulado la proliferación indiscriminada de instrumentos y modos de evaluación, aunque no siempre los productos resultantes se corresponden con los constructos teóricos para los que fueron desarrollados (Schumm, 1982). Este problema, que se concreta en la desconexión entre las formulaciones conceptuales y las estrategias de medición, se hace patente en la frecuente separación entre los profesionales dedicados al desarrollo de modelos y propuestas teóricas sobre el ajuste psicológico y los profesionales encargados del diseño de las técnicas que serán utilizadas para medir dicho ajuste y sus correlatos. Asimismo, la heterogeneidad en la producción de instrumentos de medida dificulta la posibilidad de establecer una clasificación adecuada de los mismos que permita comprender sus características constitutivas y oriente de forma adecuada a los profesionales para que tomen una decisión en cuanto a su uso.

La tercera gran dificultad guarda relación con la desconexión entre los procedimientos de medida y las estrategias de combinación de la información, de análisis de datos y de interpretación de los resultados. En este sentido, por ejemplo, sabemos que los programas de ordenador son capaces de manejar cualquier tipo de dato, proceda de un instrumento bien o mal construido, pero no son capaces de discriminar en qué grado la información que se maneja es fiable.

En la parte empírica de este proyecto de investigación presentamos una herramienta informatizada para la evaluación psicodiagnóstica de niños y adolescentes (una entrevista diagnóstica estructurada), cuyo referente teórico se encuentra en el denominado modelo médico-categorial. En el diseño del instrumento se han contemplado aspectos referidos a la gestión de datos que han demostrado estar relacionados con la calidad de los resultados obtenidos en el proceso de evaluación psicodiagnóstica.

1.2. LAS DOS GRANDES ORIENTACIONES DIAGNÓSTICAS QUE HAN GUIADO EL PROCESO DE CLASIFICACIÓN EN PSICOPATOLOGÍA

Durante años ha quedado claramente establecida la necesidad de disponer de algún tipo de clasificación para los trastornos mentales, aunque no siempre ha habido un claro consenso sobre qué enfermedades incluir y qué procedimiento utilizar para organizarlas. Esto es, a todos los problemas típicos de la evaluación, que por razones obvias no serán discutidos en este trabajo, se añade la dificultad del "etiquetado diagnóstico".

Es evidente que el diagnóstico sirve para clasificar a los sujetos dentro de un determinado grupo en función de las características que presentan. No en vano se considera que la clasificación constituye una de las tareas fundamentales del diagnóstico clínico. En la actualidad se acepta que cualquiera de los modelos teóricos existentes en el seno de la psicopatología (médico, comportamental, psicodinámico, sociológico, matemático, etc.) participa en mayor o menor grado de alguno de estos dos grandes enfoques clasificatorios: el diagnóstico clínico categorial y el diagnóstico estadístico o dimensional (Grayson, 1987).

La orientación categorial deriva de la tradición médica, de modo que utiliza criterios subjetivos consensuados. La orientación dimensional es más propia de la psicología, y se basa en categorías empíricas derivadas (fundamentalmente) del análisis factorial. Curiosamente, y a diferencia de lo que se pudiera suponer, en psicopatología la clasificación clínica categorial ha constituido el marco gnoseológico principal y ha determinado en gran medida todo el posterior proceso de evaluación, tal vez como consecuencia del grado de robustez que se han atribuido a las medidas categóricas básicas utilizadas para designar aspectos comportamentales humanos (Millon, 1991).

1.2.1. El diagnóstico como sistema de clasificación categorial

Un sistema de clasificación constituye una forma de describir y etiquetar los fenómenos de forma sistemática, utilizando para ello categorías o dimensiones comportamentales. En este apartado presentamos los principales sistemas de clasificación categoriales existentes en el seno de la psicopatología. La característica básica de estos sistemas consiste en la clasificación de los problemas del comportamiento a partir de agrupaciones discretas.

Como hemos indicado anteriormente, el diagnóstico categorial procede del ámbito de la medicina, y por consiguiente se fundamenta en el denominado modelo biológico, biofísico u orgánico. En particular, el psicodiagnóstico categorial procede del campo médico psiquiátrico (Guze, 1989), en el cual predominan los criterios clasificatorios etiopatogénicos, anatomopatológicos y/o sintomáticos. En función de estos criterios se establecen los diferentes sistemas de clasificación, que difieren entre sí en el número de categorías que incluyen y en el énfasis sobre la fenomenología, etiología y curso como rasgos definitorios.

El modelo médico sostiene dos enunciados básicos: (1) la persona con trastornos de conducta está enferma, y (2) la enfermedad del paciente se expresa a través de un síndrome característico, formado por la agregación de signos y síntomas psicopatológicos. A su vez, la modelización del sistema médico categorial se basa en un conjunto de postulados básicos (Vallejo, 1998): (a) la enfermedad tiene una etiología o causa que la diferencia claramente del estado de salud, (b) la etiología es de naturaleza orgánica, (c) la etiología orgánica produce unos síntomas que constituyen el cuadro clínico, (d) a partir de la agrupación sistemática de los síntomas es posible determinar un proceso de diagnóstico, (e) a través del diagnóstico es posible establecer un criterio pronóstico, y (f) el tratamiento (habitualmente de naturaleza biológica) incide sobre la etiología orgánica de la enfermedad.

Como han venido señalando diferentes autores (Quintanilla, 1986; Vallejo, 1998), parece que la capacidad explicativa del modelo médico, cuando se aplica de forma estricta en el ámbito de la psicopatología, es cuestionable desde cualquiera de los postulados básicos enunciados. Así, por ejemplo, los criterios que defienden sustratos etiológicos y anatomopatológicos conllevan serias dificultades, ya que el número de trastornos con etiología o anatomopatología conocida no permiten definir toda una nosografía psiquiátrica (en nuestra disciplina es habitual desconocer todas las causas, o bien considerar que éstas conllevan la aparición de problemas de muy diversa índole). Por esta razón prevalecen las clasificaciones basadas en una sintomatología descriptiva, y los sistemas de clasificación derivados de la clínica se construyen a partir del consenso de los clínicos sobre el hecho de que determinadas características se presentan conjuntamente.

Y es precisamente aquí donde nos encontramos una de las críticas más básicas de las taxonomías basadas en síntomas: la dificultad para considerarlas como "científicas" o totalmente válidas, precisamente porque están basadas en elementos externos y no etiológicos (Kendell, 1989). Pensemos que una posición meramente descriptiva conlleva necesariamente una divergencia diagnóstica poco fiable, donde cada experto centrará su atención en unos elementos concretos en función de su formación, el nivel de entrenamiento, las necesidades del caso, etc.

Desde una perspectiva netamente histórica, hasta la década de los 60 la posibilidad de obtener clasificaciones consistentes no figuraba como centro de interés de clínicos e investigadores, los cuales atribuían más relevancia al estudio de los componentes etiológicos que al estudio de los componentes descriptivos y clasificatorios (Reich, 1992). Como resultado de esta situación hallamos diversos trabajos de investigación, entre ellos la conocida crítica de Meehl (1954) sobre la fiabilidad del juicio clínico frente al estadístico, que denuncian el hecho de que gran parte de los psicólogos y de los psiquiatras abordan el proceso de medición y de diagnóstico de una forma ciertamente subjetiva y poco rigurosa, aplicando criterios de diagnóstico en ausencia de sistematización y operacionalización.

Una primera solución a este problema consistió en el análisis de diversas descripciones y el posterior nacimiento de criterios clasificatorios comunes, tal como ilustran desde hace varias décadas la RDC ("Research Diagnostic Criteria"), el DSM ("Manual diagnóstico y estadístico de los trastornos mentales") o la CIE ("Clasificación internacional de enfermedades"). Se trata de sistemas de clasificación que describen detalladamente cada trastorno en base a diferentes elementos: sintomatología básica y asociada, edad de inicio, curso, complicaciones, deterioro, factores predisponentes, datos de prevalencia, incidencia en cada sexo, antecedentes familiares, diagnóstico diferencial, criterios para el diagnóstico, etc. Los DSM han sido, con mucho, los sistemas de clasificación dominantes en Estados Unidos, y también han sido ampliamente utilizados en Europa. Esto ha originado un gran volumen de bibliografía sobre su desarrollo, y también propuestas metodológicas para analizar su validez (Frances, Widiger y Pincus, 1989; Frances et al., 1990, 1991; Kendell, 1989; 1991; Lahey et al., 1990; Nelson-Gray, 1991; Spitzer y Williams, 1988; Widiger et al., 1991).

Otro aspecto del proceso de desarrollo de estas taxonomías tiene que ver con la población para la cual han sido elaboradas. Durante décadas, los trastornos infantiles no gozaron de una consideración especial en los sistemas nosológicos y en las taxonomías vigentes, de tal forma que las clasificaciones del comportamiento "anormal" se centraron principalmente en el comportamiento de los adultos. Para justificar esta falta de atención se argumentaron diferentes razones (Kazdin, 1988): que muchas de las conductas que caracterizan los procesos de alteración o desajuste emocional son relativamente frecuentes en población infantil (sin que esta presencia requiera de una atención diagnóstica particular), que la infancia es una etapa caracterizada por múltiples y rápidos cambios que dificultan la aprehensión de estas alteraciones, o incluso que es habitual que el desarrollo infantil conlleve la atenuación de muchas conductas problemáticas. En la actualidad, sin embargo, tanto clínicos como investigadores reconocen la importancia de considerar la psicopatología infantil como un cuerpo de conocimientos particular diferenciado de la psicopatología de adultos. Esto ha hecho que las taxonomías más recientes disponibles en nuestra disciplina contemplen procesos considerados característicos de la población infantil y juvenil, o bien que describan las formas específicas en que las distintas alteraciones se manifiestan a estas edades.

Los sistemas categoriales aceptan unas premisas básicas para la clasificación de las conductas problema (Werry, 1992):

1. Los síntomas que forman cada trastorno deben estar perfectamente descritos y definidos operacionalmente.
2. Los elementos pertenecientes a una categoría (en este caso, grupo diagnóstico) deben ser semejantes entre sí y claramente diferenciables de los elementos de otras clases.
3. Deben existir límites claros y fácilmente objetivables entre las categorías.
4. Las características que definen cada clase han de permanecer constantes.
5. La colocación de un sujeto en una categoría debe ser estable y permanente.
6. Las diferentes clases deben estar definidas de modo que resulten mutuamente excluyentes entre sí.
7. Finalmente, el sistema de clasificación en conjunto debe demostrar ser específico, fiable, válido y útil.

Sin embargo, todos y cada uno de estos principios es cuestionable en todas las clasificaciones categoriales de que disponemos en la actualidad, incluso en las más utilizadas. Así queda enunciado, por ejemplo, en las críticas que se incluyen en el propio manual del DSM-IV (p. XXII, versión castellana, 1995). Por esta razón, creemos conveniente señalar que los sistemas de clasificación categorial no están exentos de deficiencias, y que a lo largo de las dos últimas décadas se les han imputado numerosas críticas. Por ejemplo, algunos autores destacan que las clasificaciones en conjunto resultan más "médicas" que psicológicas, y que la

postura "ateórica" que ha guiado su elaboración también ha dificultado el desarrollo de una base empírica adecuada para demostrar las diferencias cualitativas entre las diferentes clases, ha disminuido la validez de los cuadros indicados y ha conllevado a un excesivo grado de inexactitud para las valoraciones de algunos síntomas (Kazdin, 1989). Además, el hecho de basarse en clases o grupos de inclusión-exclusión disminuye la posibilidad de utilizar importantes medidas psicológicas, como la dimensionalidad y/o el gradiente de intensidad con el que se presentan las conductas (Fergusson y Horwood, 1995). Finalmente, también se ha considerado que el abordaje categorial es poco sensible para medir los cambios en el estado psicopatológico de los sujetos (especialmente los pequeños), ya que requiere que transcurra un tiempo considerable para que los trastornos cambien su estado diagnóstico (Boyle et al., 1996).

A pesar de todas las deficiencias expuestas, estas clasificaciones han sido ampliamente utilizadas y han originado una ingente investigación clínica y epidemiológica. Por ejemplo, los DSM, cuyos criterios diagnósticos descriptivos pretenden ser explícitamente ateóricos, constituyen una herramienta de clasificación aceptada internacionalmente. Muchos de los trabajos publicados en las revistas más prestigiosas de nuestra ciencia utilizan este manual, y esto a pesar de otras numerosas pruebas existentes respecto a la influencia de factores étnicos y culturales en la aparición y curso de un gran número de trastornos. Tal vez por esta razón durante el proceso de preparación de la última versión (DSM-IV) se ha procurado que el manual pueda ser utilizado en poblaciones de diferentes culturas a la estadounidense. La discusión de en qué medida se haya alcanzado este propósito no constituye el objeto de este trabajo. En principio, lo que sí parece claro es que el DSM-IV es compatible con la Clasificación Estadística Internacional de Enfermedades y Problemas de Salud en su décima revisión (CIE-10), llevada a cabo por la OMS y publicada en 1992 (W.H.O, 1992).

En fin, los sistemas taxonómicos como los DSM gozan de un uso generalizado, y representan un ejemplo de la necesidad existente en el ámbito de la psicopatología para utilizar sistemas criterios de diagnóstico y de fijar árboles de decisión excluyentes y diferenciales (Zarin y Earls, 1993). El valor y la utilidad práctica de estos sistemas es sobradamente conocido por la mayoría de los profesionales. Por este motivo, es deseable que en un futuro próximo se les dote de mayor operatividad y rigor metodológico y científico.

1.2.1.1. Evaluación diagnóstica categorial a través de entrevistas estructuradas

El procedimiento de evaluación más habitual dentro del diagnóstico categorial es la entrevista. Existen diferentes tipos de entrevistas, entre las que destacan: las clínicas (también denominadas biográficas) y las diagnósticas (Wiens, 1990). Las primeras están especialmente diseñadas para recabar información sobre el desarrollo y el ambiente psicológico y social, es decir, la historia del sujeto. Las segundas están especialmente diseñadas para establecer el diagnóstico (Dohrenwend, 1998), y por esta razón les vamos a dedicar un breve espacio.

En una primera fase, la evaluación psicodiagnóstica de los sujetos se realizó fundamentalmente a través de entrevistas clínicas abiertas, ya que existía un alto grado de acuerdo a la hora de considerar estos instrumentos como los métodos de evaluación más completos sobre la conducta, las emociones y las cogniciones del sujeto; a través de la información recogida en estas entrevistas, el clínico establecía el diagnóstico adecuado según la taxonomía que utilizara y buscaba una explicación etiológica y funcional del problema. Pero pronto se hizo evidente la existencia de numerosas fuentes de variabilidad que incidían en este tipo de entrevista: la variabilidad de la información (o lo que es igual, la forma de buscar y obtener la información de los individuos), la variabilidad del sujeto (o variación en el contenido y en el estilo de contestar de un informante a otro), la variabilidad de criterio (debida al uso de diferentes sistemas/criterios de clasificación diagnóstica), la variabilidad de

la situación y la variabilidad de la observación (Ezpeleta, 1996; Fowler y Mangione, 1990; Gutterman, O'Brien y Young, 1987). Estos estudios han originado numerosas críticas frente a la entrevista clínica abierta, fundamentalmente como consecuencia del componente subjetivo que puede dictar tanto la recogida como la interpretación de la información. Este factor ha hecho que la fiabilidad del diagnóstico entre evaluadores haya sido moderada o baja en múltiples investigaciones, tanto con adultos como con niños (Anthony et al., 1985; Helzer et al., 1985; Kwoh et al., 1992; Rey, Plapp y Stewart, 1989; Robins, 1985; Vitiello, Malone, Buschle, Delaney y Behar, 1990; Weinstein, Stone, Noam, Grimes y Schwab-Stone, 1989). Esto no debe sorprendernos si tenemos en cuenta que cuando no existe un procedimiento reglado, los diferentes clínicos suelen centrarse en problemas concretos diferentes, (ya que tienen distintas concepciones de la dinámica de los trastornos), o bien diagnostican usando fuentes de información distintas.

Debido a esta falta de acuerdo entre clínicos, los investigadores se han visto en la necesidad de desarrollar un sistema estandarizado para obtener la información. Y es aquí donde nos encontramos el segundo gran cambio en la década de los 80: el desarrollo de entrevistas estructuradas, que poco a poco ganan protagonismo sobre los tradicionales métodos de entrevista abierta (Ezpeleta, 1996; Silverman y Kearny, 1992; Vitiello et al., 1990). Las entrevistas diagnósticas estructuradas también fueron creadas con el objetivo más limitado de obtener un diagnóstico basado en un sistema de clasificación concreta (como el DSM-IV) o para investigar grandes poblaciones a fin de comprobar la prevalencia de los trastornos.

La estructuración de una entrevista es un procedimiento que consiste en la definición precisa del fenómeno a valorar, especificando el orden y la redacción concreta de las preguntas, la estandarización de la forma de valoración, registro, combinación e interpretación de las respuestas. Además de otras ventajas, es importante tener presente que este proceso limita la variabilidad en el intercambio de preguntas y respuestas entre el entrevistador y el entrevistado (Beitchman, Kruidenier, Clegg, Hood y Corradini, 1989; Ezpeleta, 1995; Gutterman et al., 1987).

En función del grado de rigidez en que se especifiquen las preguntas, permitan hacer un diagnóstico o se estandarice la administración, se puede hablar de entrevistas estructuradas o semiestructuradas. Las entrevistas estructuradas se caracterizan por especificar exactamente la formulación, la secuencia de las preguntas y las normas para evaluar y registrar las respuestas del sujeto; esto hace que el juicio subjetivo del entrevistador esté minimizado ya que se le da poca libertad para dirigir la entrevista. De hecho con estos protocolos no es necesario que el entrevistador tenga experiencia clínica y pueden utilizarlos personas en formación con una cualificación general (al menos teóricamente), aunque no sean especialistas y carezcan de experiencia clínica (entrevistador lego): el entrevistador constituye un elemento intercambiable del proceso de evaluación. Aunque se ha acusado a las entrevistas diagnósticas estructuradas de ser rígidas, mecánicas y no permitir adaptarse a un informante particular, cuentan con la innegable ventaja de que permiten obtener datos objetivos y fiables (Ezpeleta, 1996), lo cual ha hecho que estos protocolos se utilicen frecuentemente en la investigación. Las entrevistas semiestructuradas, sin embargo, se caracterizan por no ser tan rígidas y dar al entrevistador más flexibilidad para dirigirlas, por lo cual éste requiere necesariamente de experiencia clínica; los protocolos semiestructurados se usan más frecuentemente en la práctica clínica. Entrevistas estructuradas y semiestructuradas tienen en común la necesidad de que los entrevistadores estén muy entrenados en el uso del instrumento de evaluación, y es también imprescindible que conozcan los criterios diagnósticos en los que se basa la entrevista (Gibbon, McDonald-Scott y Endicott, 1981).

Cualquier instrumento de evaluación debe haber demostrado con qué exactitud, eficiencia y éxito consigue su objetivo, es decir, debe haber demostrado que es fiable, válido y útil a la hora de generar el diagnóstico (Messick, 1995). Como norma general, cuanto más

estructurada sea la entrevista y mejor definidos estén los procedimientos de evaluación y corrección del protocolo, se obtendrá un mayor acuerdo entre los entrevistadores. La importancia del estudio de las propiedades psicométricas de los instrumentos de evaluación ha conducido a una gran diversidad de estudios metodológicos orientados al análisis de la fiabilidad (fiabilidad entre observadores, fiabilidad test-retest, etc.) y validez de las entrevistas diagnósticas, tanto en el ámbito de la epidemiología como de la clínica. El centro de interés de dichos estudios de fiabilidad ha sido la estabilidad de la escala en su conjunto o del resultado final obtenido a partir de la escala. Además, a través de los estudios de fiabilidad y validez de los instrumentos de evaluación se obtiene información de la estabilidad temporal de un determinado diagnóstico, lo cual constituye un rasgo esencial de la validez de dicho trastorno, ofrece la base sobre la cual predecir el curso y desarrollo, proporciona conocimiento de la semiología de la enfermedad y, en última instancia, permite modificar y mejorar el propio sistema taxonómico. Asimismo, el estudio de la fiabilidad y estabilidad de los diagnósticos en la infancia y en la adolescencia proporciona una oportunidad para evaluar el efecto del estadio de desarrollo del niño en la aparición y evolución de la psicopatología infanto-juvenil (Mattanah et al., 1995).

Se asume que todas las entrevistas estructuradas y semiestructuradas más utilizadas en psicopatología poseen índices adecuados de fiabilidad, y también de validez de contenido, puesto que siguen los criterios de clasificación de la APA (DSM-III, DSM-III-R o DSM-IV) y en ellas se incluye una selección de ítems representativa de cada categoría (Ezpeleta, 1990, 1995). Sin embargo, nos gustaría señalar que el estudio y la interpretación de las propiedades psicométricas es especialmente difícil en el caso de entrevistas diagnósticas estructuradas, ya que la validación está inevitablemente afectada por la validez del sistema diagnóstico en el que se basan (por ejemplo el DSM).

El reconocimiento de las ventajas que supone el uso de entrevistas diagnósticas estructuradas ha supuesto un cambio notable en las actitudes de la mayoría de psicólogos y psiquiatras que las utilizan, los cuales consideran que la demostración de fiabilidad y validez aseguran la calidad de los datos que se registran, aportando la solución que requerían los problemas de medición en psicopatología. Sin embargo, la experiencia acumulada a lo largo de los últimos años ha demostrado que la realidad es bien diferente. Si bien es cierto que el desarrollo de las entrevistas estructuradas ha supuesto un importante avance en el campo de la evaluación psicológica y del psicodiagnóstico, todavía queda un gran camino por recorrer en lo que se refiere a mejoras en validez y calidad de datos (Rutter y Pickles, 1990). De hecho, es habitual que numerosos clínicos e investigadores que utilizan estos instrumentos para la recogida de información acepten "a priori" como válidos los resultados directos obtenidos en la evaluación, y que por ello no presten la suficiente atención a los procedimientos específicos que utilizan en los procesos de medición, manejo de datos y definición de los correspondientes diagnósticos (Zarin y Earls, 1993). En este sentido, consideramos que aunque los profesionales posean herramientas muy potentes para la evaluación de los trastornos psicológicos, en el mejor de los casos no obtienen de ellas el mejor rendimiento, y en el peor de los casos (o lo que es igual, la mayoría de las veces) las utilizan mal.

Somos conscientes de que el grado de validez y la calidad de los resultados obtenidos en el ámbito de la evaluación dependen de la consideración de múltiples factores metodológicos y procedimentales. En este sentido consideramos que en la actualidad persisten, al menos, tres importantes obstáculos relacionados con el uso de las entrevistas estructuradas: el diseño de protocolos con garantías de estandarización, la disponibilidad de herramientas que sean fácilmente aplicables a diferentes poblaciones de sujetos y la garantía de calidad de los procesos de gestión y manejo de datos. Por esta razón, en la parte empírica de este trabajo presentamos el proceso que hemos seguido para informatizar una entrevista estructurada adecuada para la evaluación psicodiagnóstica en población infantil y juvenil. Uno de los objetivos principales de la adaptación de esta entrevista al nuevo formato informatizado ha

consistido, precisamente, en mejorar todos aquellos aspectos relacionados con la medición, en especial los procesos de estandarización de la herramienta y gestión de datos clínicos. Estas mejoras suponen un ahorro importante en el tiempo de administración, procesamiento y gestión de datos, a la vez que contribuyen a incrementar notablemente la calidad de los resultados de nuestra investigación.

1.2.1.2. Entrevistas diagnósticas estructuradas en población infantil y juvenil

Durante los últimos años, sensibilizados por el problema de la evaluación infantil y por la problemática de la calidad de los datos, hemos estado trabajando en la adaptación de una entrevista estructurada para la evaluación psicodiagnóstica en población infantil y juvenil. Además de estudiar las características psicométricas del protocolo en el formato habitual de lápiz y papel, hemos elaborado una nueva versión en formato informatizado.

Previamente a la presentación del protocolo escogido por nosotros, destacaremos algunas de las características que habitualmente se han relacionado con el uso de entrevistas diagnósticas estructuradas cuando se está evaluando a niños pequeños. La razón de este preámbulo estriba en que el trabajo con niños pequeños es sustancialmente diferente del trabajo con adultos. Por esta razón, en los siguientes párrafos se resumen los que, a nuestro juicio, son los factores fundamentales relacionados con este tipo de evaluación.

En la actualidad se acepta que el proceso de evaluación de niños y adolescentes comparte las características generales y comunes a toda evaluación, si bien diferentes autores matizan que en población infanto-juvenil este proceso conlleva unos aspectos más o menos específicos (Achenbach, McConaughy y Howell, 1987; Del Barrio, 1996; Forns, 1993; Herjanic, 1984; Silva, 1995): se deben considerar la perspectiva evolutiva y las variables ambientales del proceso, y se debe realizar una evaluación multicomponente (diferentes tipos de instrumentos de evaluación), con múltiples informantes (el propio sujeto, los cuidadores, los profesores o los compañeros) y en diferentes áreas (familia, escuela, etc.).

Pero esto no siempre fue así. Durante mucho tiempo la evaluación psicológica del niño se ha centrado fundamentalmente en la entrevista clínica con los padres. Esto parecía hasta cierto punto razonable, ya que se asumía que los padres son quienes poseen más información referida al desarrollo, comportamiento y habilidades de los hijos (Loeber, Green y Lahey, 1990; Loeber, Green, Lahey, y Stouthamer-Loeber, 1989, 1991; Williams, McGee, Anderson y Silva, 1989). Muchos investigadores llegaron incluso a cuestionar la validez de los niños como informantes fiables de su propio estado cuando son interrogados de forma directa, argumentando que los más pequeños tienen serias dificultades para mantener la atención durante el proceso de evaluación, o bien que no poseen las habilidades suficientes para expresar sus propios comportamientos y sentimientos, o para referir síntomas psicopatológicos (Herjanic, Herjanic, Brown y Wheatt, 1975; Schwab-Stone, Fallon, Briggs, y Crowther, 1994). Por ello se desarrollaron numerosas técnicas para obtener información de los niños más pequeños de forma indirecta (juegos, dibujos, títeres, etc.). Estas técnicas dejaban jugar y asociar al niño, y obligaban al examinador a interpretar el significado de las respuestas obtenidas a partir de su propio juicio como profesional.

Sin embargo, en las últimas décadas se han producido notables cambios en el ámbito de la psicopatología y del psicodiagnóstico infantil. Tal vez la más significativa sea el incipiente debate sobre la capacidad de comprensión y participación de los niños en el proceso de evaluación y diagnóstico (Edelbrock y Costello, 1990; Edelbrock, Costello, Dulcan, Conover y Kalas, 1986; Helzer, 1983; Herjanic, 1984). En efecto, en los últimos años ha crecido la convicción de que los niños son por sí mismos informantes consistentes y valiosos de sus propios comportamientos y sentimientos (Ezpeleta, 1990; Lewis, Gorsky, Cohen y Hartmark, 1985; Zahner, 1991). Cada vez más investigadores clínicos coinciden en que los padres no

proporcionan necesariamente información más válida y fiable que otros informantes (como, por ejemplo, los profesores o el propio sujeto), aunque también es cierto que las percepciones paternas continúan siendo cruciales en el proceso de evaluación y tratamiento (Ezpeleta, 1990; Gutterman et al., 1987; Herjanic, 1984; Klein, 1991). En cualquier caso, en la actualidad se acepta que el trabajo con niños y adolescentes debe reunir diferentes características: implicar a múltiples personas (niños, padres y profesores, fundamentalmente), utilizar diferentes métodos de evaluación (entrevistas, observación directa, autoinformes, etc.), tener en cuenta el nivel de desarrollo del sujeto y obtener información sobre la capacidad cognitiva del niño (La Greca y Stone, 1992). Este cambio de orientación ha hecho necesario el desarrollo de instrumentos de evaluación directa con el niño, tanto en el ámbito de la práctica clínica como en el ámbito de la investigación (Herjanic, 1984; Young, O'Brien, Gutterman y Cohen, 1987a, 1987b; Werry, 1992).

Entre los elementos clave que facilitaron el nacimiento de las entrevistas diagnósticas estructuradas (y también semiestructuradas) para niños y adolescentes cabe subrayar la aparición de diferentes sistemas de clasificación diagnóstica con categorías específicas para niños y adolescentes, lo cual hizo necesarios procedimientos reglados para evaluar estas alteraciones y el grado en el que se presentan. Afortunadamente, en la actualidad está ampliamente reconocida y aceptada la necesidad de utilizar técnicas de entrevistas estandarizadas si se desean obtener diagnósticos fiables, tanto en el ámbito de la investigación como de la práctica clínica habitual (Harrington et al. 1988). En lo que se refiere a los protocolos generales para el diagnóstico en niños y adolescentes que cubren un amplio espectro de psicopatología, destacan la Diagnostic Interview for Children and Adolescents (DICA; Herjanic et al., 1975; Reich, Shayka y Taibleson, 1991; Washington University, 1983) y la Diagnostic Interview Schedule for Children (DISC; Costello, Edelbrock, Kalas, Kessler y Klaric, 1982; Ribera et al., 1996; Roberts, Solovitz, Chen y Casat, 1996; Shaffer, 1994; Shaffer et al., 1983, 1993, 1996; Schwab-Stone et al., 1993) como entrevistas estructuradas, y la Child Assessment Schedule (CAS; Hodges, Kline, Fitch, McKnew y Cytryn, 1981), la Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS; Ambrosini, 1988; Ovarschel, Puig-Antich, Chambers, Tabrizi y Johnson, 1982; Puig-Antich y Chambers, 1978; 1983) y la Child and Adolescent Psychiatric Assessment (CAPA; Angold y Costello, 1995; Angold, Cox, Prendergast, Rutter y Simonoff, 1993) como entrevistas semiestructuradas. Estas entrevistas difieren en el entrenamiento que requieren los entrevistadores, la duración, las categorías diagnósticas cubiertas, las edades para las que se han diseñado y la organización (una descripción detallada se encuentra en Edelbrock y Costello, 1990; Ezpeleta, 1996; Reich, 1992). El protocolo escogido por nosotros ha sido la Diagnostic Interview for Children and Adolescents, DICA, cuyas características se han resumido en el capítulo 5.

1.2.2. El diagnóstico dimensional

Hemos indicado que el diagnóstico psicopatológico propuesto por sistemas como el DSM se organiza en torno a categorías discretas. Este tipo de clasificación, propia de la ciencia médica, implica que una persona determinada es asignada a una categoría diagnóstica cuando comparte los criterios definidos de inclusión y/o exclusión a la clase.

A diferencia de los sistemas categoriales, los dimensionales ofrecen una clasificación basada en la cuantificación de los atributos. Se trata de modelos de tradición psicométrico-estadística, caracterizados por buscar la "relación matemática" subyacente a los distintos fenómenos que se desean explicar (y no por establecer presupuestos sobre las relaciones observadas). Uno de los modelos matemáticos que ha alcanzado mayor difusión en el seno de la psicología y que se encuentra en la base de gran parte de los modelos dimensionales del

comportamiento es el factorial, caracterizado por la posibilidad de medir el grado en que los atributos se presentan en los sujetos para poder situarlos a lo largo de categorías continuas (McDonald, 1985). Además de este modelo, que permite analizar los diferentes agrupamientos empíricos, también se han utilizado otras técnicas como el análisis de "clusters" o el análisis discriminante, que permiten definir diferentes grupos de diagnóstico (Fergusson y Horwood, 1995; Hasselblad y Hedges, 1995).

Desde una perspectiva funcional, los sistemas dimensionales priorizan la medición de los cambios conductuales, entendidos como variables continuas. La forma en que operan estos sistemas consiste en dar valores cuantitativos que indican el grado de intensidad de las características que presentan los sujetos en diversas dimensiones. El conjunto de estas dimensiones determina un perfil específico en cada individuo. Los trastornos son una consecuencia de la magnitud y/o combinación empírica de continuos y perfiles (Fergusson y Horwood, 1995). Esta consideración posee dos repercusiones teóricas muy significativas que diferencian los sistemas dimensionales de los categoriales: (1) la "normalidad" y la "anormalidad" no constituyen fenómenos distintos, sino que forman parte del mismo continuo; y (2) todos los comportamientos participan del rango normal.

Uno de los inconvenientes más relevantes de este enfoque consiste en la dificultad para decidir y definir cuáles son las dimensiones matemáticas relacionadas con los fenómenos psicopatológicos que se deben medir, ya que las que se observan no siempre tienen por qué ser las más relevantes o las más válidas (Millon, 1991). Además, una vez que se obtienen los datos y las diferentes dimensiones, este modelo necesita explicarlos recurriendo a conceptos de base biológica, comportamental, sociológica o psicodinámica.

La tradición clínico-psiquiátrica y la dimensional han coexistido de forma no sencilla. En la última década se ha exacerbado la oposición entre ambos abordajes en el ámbito del psicodiagnóstico (Kazdin, 1988, 1989; McReynolds, 1989; Trull, Widiger y Guthrie, 1990; Widiger y Trull, 1991), en parte como consecuencia del incremento de las críticas al modelo categorial. Si intentamos sintetizar las conclusiones más relevantes de este debate destacaríamos la constancia de que el uso exclusivo de sistemas de clasificación categorial presenta dificultades teóricas y prácticas. Por ejemplo, la delimitación de los puntos de corte que diferencian el límite entre la conducta normal y la patológica, o la dificultad para establecer una clara diferenciación de conductas cuando éstas se comportan como elementos de conjuntos borrosos.

En efecto, numerosos estudios constatan que en muchas ocasiones el diagnóstico dimensional constituye una postura más ventajosa y útil que la categorial a la hora de describir procesos que no poseen límites definidos (sobre todo porque su distribución sigue una función continua), y también en el momento de definir y detectar complejos que podrían pasar desapercibidos en los sistemas categoriales (Blashfield y Livesley, 1991; Costello, 1992; Millon, 1991; Morey, 1991). En relación a este último punto, es importante señalar que cuando se dispone de un perfil de comportamientos es posible describir a nivel real los casos poco frecuentes, los cuales suelen ser excluidos de las tipologías clásicas debido precisamente a que dichas condiciones infrecuentes no son consistentes con las categorías prescritas (Ávila, Rodríguez-Sutil y Herrero Sánchez, 1996).

Entre otras ventajas importantes relacionadas con el uso de sistemas psicométricos, también se ha destacado la flexibilidad que ofrecen los diagnósticos dimensionales, ya que posibilitan definir diferentes puntos de corte en función de los objetivos previos (predicción, prevención, tratamiento, etc.). Además, la perspectiva dimensional supone una pérdida menor de información potencialmente importante, a la vez que concede a las diferentes características comportamentales la misma atención (cosa que no sucede cuando una única característica distintiva aparece en primer plano en una tipología categorial).

Finalmente, diferentes estudios de investigación constatan que las medidas dimensionales poseen mejor validez predictiva que las basadas en criterios tales como los

DSM (Fergusson y Horwood, 1995). Esta afirmación se basa en diferentes evidencias. Por ejemplo, numerosos trastornos (por ejemplo los de carácter externalizante) han demostrado que están formados de síntomas y signos cuya estructura queda mejor explicada a través del modelado dimensional (Boyle et al., 1996; Fergusson y Horwood, 1995; Jensen et al., 1996; Rubio-Stipec et al., 1996). Asimismo, la dicotomización de los trastornos en dos categorías (presente "versus" ausente) según la correspondiente distribución de síntomas implica una sustancial pérdida de información, la cual sería una de las causas fundamentales de la notable pérdida de poder predictivo. En cualquier caso, al hablar de validez, no podemos olvidar que la validez predictiva de los métodos depende del contexto donde se apliquen. Así, el uso de procedimientos de clasificación categorial en población general es especialmente cuestionable, ya que conduce a distribuciones muy asimétricas en las cuales la mayoría de sujetos se clasifica como "no casos" y sólo una pequeña parte de los individuos forman parte del grupo de "casos". Y como cualquier clínico sabe, los sujetos considerados "clínicamente sanos" representan un grupo muy heterogéneo en lo que se refiere a la configuración de síntomas y el riesgo de psicopatología. También creemos importante matizar que los sistemas dimensionales parecen ser más coherentes con la realidad que se observa en la práctica clínica, ya que fomentan y facilitan la representación del carácter idiosincrásico de los sujetos en lugar de "forzarlos" a encajar en unas categorías definidas "a priori" (Haynes y O'Brien, 1990).

Pero a pesar de las ventajas demostradas por los modelos dimensionales, en la comunidad científica todavía se continúan utilizando ampliamente los sistemas categoriales como herramienta de comunicación a la hora de transmitir la información. Muchos clínicos e investigadores aluden a la necesidad de acaparar más datos referentes a los sistemas dimensionales, ya que debido al menor uso realizado de estos modelos, aún no existe acuerdo sobre las dimensiones que deben utilizarse para clasificar o de cuáles son menos aceptadas. Además, el diagnóstico dimensional, precisamente por su naturaleza cuantitativa, requiere de grandes poblaciones para definir los datos normativos y disponer de la suficiente base estadística como para generar los correspondientes "clusters" que constituirán los distintos trastornos.

1.2.2.1. Instrumentos de evaluación diagnóstica dimensional

Dentro del diagnóstico dimensional destaca la aparición de múltiples instrumentos diseñados para ser administrados a grandes poblaciones, como inventarios y cuestionarios que se aplican tanto en la población adulta como en la infantil y juvenil (Hasselblad y Hedges, 1995; Pedreira y Sánchez-Gimeno, 1992). Estos instrumentos se organizan en función del tipo de trastorno que evalúan (unidimensionales), aunque también existen algunos "listados globales" que miden cualquier tipo de alteración (pluridimensionales). La mayor parte se han desarrollado como autoinformes, y se caracterizan por recabar información verbal sobre los sujetos, sus comportamientos y las situaciones en las que tienen lugar las conductas (Fernández-Ballesteros, 1992).

En cuanto a las estrategias de construcción, la mayoría de los instrumentos de evaluación dimensional se han elaborado de acuerdo a tres estrategias diferentes: racional, empírica y factorial (Hasselblad y Hedges, 1995; Millon, 1991). En el primer caso, los instrumentos "racionales" están formados por reactivos (preguntas, ítems) seleccionados por los evaluadores (presumiblemente "expertos") en base a un modelo teórico inicial según el interés de su contenido para medir una determinada variable o aspecto del comportamiento. Los instrumentos construidos de forma empírica incluyen reactivos que han demostrado previamente su capacidad para discriminar entre grupos de sujetos. Finalmente, los instrumentos construidos utilizando la estrategia factorial seleccionan los ítems que obtienen los mejores pesos factoriales en los análisis realizados. Queremos destacar también que es

habitual que los instrumentos de evaluación dimensional de última generación combinen diferentes estrategias de creación con objeto de dotarlos de una mayor calidad.

En cuanto a las áreas de evaluación, se han construido diferentes instrumentos según las variables que se desean medir y el modelo teórico de referencia (Miguel-Tobal, 1995). Así, por ejemplo, desde el enfoque de rasgos, factores o dimensiones de personalidad, ha sido habitual diseñar procedimientos psicométricos (tales como cuestionarios) para evaluar las variables "intrapersonales" (supuestamente estables) que se toman como factores explicativos de los comportamientos. Desde el enfoque conductual, los instrumentos de evaluación, que se han diseñado utilizando estrategias preferentemente racionales, han prestado una mayor atención a aspectos de tipo situacional, entendiendo las respuestas de los sujetos como muestras de conducta (y no como disposiciones internas). Finalmente, bajo el enfoque cognitivo y cognitivo-conductual, se han construido cuestionarios sobre repertorios, procesos y estructuras cognitivas.

1.3. UNA VISIÓN DE CONJUNTO

Bajo nuestro punto de vista, la psicopatología estudia diversos cuadros que no pueden interpretarse desde un único modelo (Yager, Bird, Staguetza-Jaramillo, Gould y Canino, 1993). Por esta razón estamos de acuerdo con algunos autores (Kazdin, 1988; Weinstein, Noam, Grimes, Stone y Schwab-Stone, 1990; Widiger, 1991) a la hora de afirmar que sería muy interesante poder aunar las taxonomías discretas y las dimensionales, por ejemplo realizar una evaluación diagnóstica categorial que ofreciera simultáneamente para cada sujeto medidas cuantitativas o utilizar técnicas comunes tales como entrevistas estructuradas junto a cuestionarios o análisis neurofisiológicos (si se desea, con diferente peso). De este modo, además de la simple asignación de los sujetos a una clase psicopatológica, se obtendrían análisis más acordes con las características individuales de los individuos de un mismo grupo, como por ejemplo medidas relativas a la severidad o contenido de las alteraciones que presentan dichos pacientes. De hecho, consideramos que los sistemas categoriales del tipo DSM no son tan ajenos al uso de dimensiones, ya que si bien es cierto que los ejes propiamente psicopatológicos carecen de la definición clara de tales dimensiones (ejes I y II, a no ser que consideremos como dimensión el número de síntomas o criterios presentes), sí que las utilizan de forma más evidente en los ejes encaminados a analizar la intensidad del estrés y la presencia de problemas psicosociales o ambientales que puedan afectar el diagnóstico, tratamiento o pronóstico (eje IV) y el nivel de adaptación y funcionamiento general de los sujetos (eje V).

En el proyecto de investigación en el que estamos trabajando durante los últimos años seguimos esta propuesta y utilizamos conjuntamente instrumentos de medida categoriales y dimensionales para valorar el estado diagnóstico de los sujetos (niños y adolescentes) y los factores que se relacionan con la presencia de psicopatología. En el capítulo 5 se presenta la estructura de la base de datos que nos permite recabar esta información.

EL ORDENADOR COMO HERRAMIENTA DE AYUDA PARA LA EVALUACIÓN PSICOLÓGICA Y LA TOMA DE DECISIONES CLÍNICAS

"Diagnoses have been shown to be more reliable once the element of interpretation is eliminated from the assessment process. (...) Simply put, interrater reliability is greater for computerized evaluations than for those conducted by therapists. (...) Basing decision on extremely formalized data is responsible for this improvement in assessment".

(Pardeck y Schulte, 1990)

2.1. INTRODUCCIÓN

En los últimos años se ha venido produciendo un importante auge en el desarrollo de tecnologías informáticas, tanto en los aspectos de "hardware" como en los de "software". La adopción y aplicación de estas nuevas tecnologías ha impulsado a su vez un notable avance en diferentes campos del conocimiento humano, entre ellos el de la psicología y el de la psicopatología.

En este segundo capítulo exponemos brevemente cuál ha sido el uso fundamental del ordenador en el ámbito general de la evaluación psicológica, y de qué modo se han utilizado los sistemas informáticos como herramientas de ayuda en el proceso de evaluación clínica y/o psicodiagnóstica. La exposición parte de los trabajos publicados tras la conocida polémica acaecida de la confrontación de las dos estrategias básicas que han orientado la toma de decisiones clínicas: la actuación clínica frente a la actuarial.

2.2. LA POLÉMICA CLÍNICO-ACTUARIAL: UNA CONFRONTACIÓN QUE ILUSTRRA LA NECESIDAD DE ADOPTAR HERRAMIENTAS DE AYUDA EN EL PROCESO DE PSICODIAGNÓSTICO

Desde hace varias décadas se ha venido cuestionando la calidad de las decisiones y juicios que emiten los profesionales en el ámbito de la salud mental. Los resultados obtenidos en numerosos trabajos de investigación sugieren que la actuación de los clínicos durante el proceso de evaluación diagnóstica no es tan buena como sería deseable (sobre todo en las tareas habituales de medición, clasificación y predicción), y esto independientemente del grado de experiencia profesional (Garb, 1989).

Estos resultados han despertado un notable interés por establecer qué factores se relacionan con las deficiencias observadas en la actuación de los clínicos que participan en trabajos de investigación (Griffin, Dunning y Ross, 1990; Griffin y Tversky, 1992). Entre las posibles variables predictoras se han señalado: el grado de seguridad que los profesionales tienden a depositar en sus propios juicios, el grado de experiencia y formación de los clínicos en tareas de evaluación psicodiagnóstica (fundamentalmente en las fases de recogida de los datos, clasificación y predicción), las propias características personales de los evaluadores, el tipo de tareas a las que se enfrentan, el tipo de pacientes que evalúan y la calidad de los instrumentos de evaluación de que disponen. A nivel general, y a diferencia de lo que muchos

podrían esperar, los resultados de estos trabajos sugieren que ni el grado de experiencia profesional, ni el entrenamiento intensivo de los futuros evaluadores, ni la cantidad de información de la que dispone el clínico, incrementan notablemente la calidad de los juicios clínicos. Bien al contrario, se ha postulado que dos de las razones básicas de la falta de calidad en la toma de decisiones clínicas han consistido precisamente en el uso de instrumentos de evaluación inadecuados y en el particular manejo que realiza el clínico de la información diagnóstica disponible (Godoy, 1996; Krol, De Bruyn y Van den Bercken, 1992).

Revisemos un poco más ampliamente este punto. En la década de los 50, y después de estudiar sin demasiado éxito qué variables podrían introducirse en el entrenamiento de los futuros profesionales en salud mental para mejorar su actuación clínica y diagnóstica, aparecieron dos trabajos muy significativos que todavía enturbiaron más el clima de negativismo. En el primero (Eysenck, 1952) quedaba claramente establecida la falta de efectividad de las psicoterapias para la mejora de los problemas psicológicos. En el segundo (Meehl, 1954) se constataba empíricamente que la actuación de los clínicos durante el proceso diagnóstico no era mucho mejor.

De hecho, el libro de Meehl (1954) acrecentó una gran polémica (originada en los años 20) en torno a la confrontación de dos tipos de predicciones, las "actuariales" (también denominadas mecánicas o estadísticas) frente a las "clínicas". De forma concisa, la predicción actuarial se definió como la asignación de enunciados interpretativos en función de la asociación estadística de dichos enunciados con los resultados de las pruebas de evaluación; esto es, se trata de una interpretación basada en la combinación matemática de la información. De otro lado, la predicción clínica es aquella que se basa en la asignación de enunciados interpretativos en función de criterios de decisión básicamente humanos; en este caso, se trata de una interpretación fundamentada en la combinación intuitiva de la información. Por supuesto, esta diferenciación no implica que los profesionales que trabajan con una estrategia clínica no utilicen datos actuariales. De hecho, cualquier clínico recurre en un momento u otro a fuentes de datos empíricos contrastados estadísticamente. Hablar de estrategia clínica denota que, en el último momento, los profesionales recurren a su propia experiencia práctica a la hora de emitir juicios e interpretaciones.

En el texto de Meehl (1954), que revisaba 20 estudios previos, se ponía de manifiesto que los modelos matemáticos eran muy superiores al método clínico (en términos de eficacia y eficiencia) de cara al establecimiento de clasificaciones y predicciones clínicas. En dicho libro también se demostraba que esta superioridad se lograba aunque los métodos estadísticos operasen sobre idénticos conjuntos de datos. Estos resultados, tomados en conjunto, reafirmaron que el problema del razonamiento diagnóstico no debía radicar tan directamente en la calidad de la información de que disponen los profesionales, sino en la forma particular en que los humanos manejan y combinan los datos clínicos.

Evidentemente, la postura de Meehl ha encontrado fuertes detractores. Uno de los mayores opositores se encuentra en Holt, quien de forma reiterada ha sostenido que el análisis de Meehl sobre el juicio clínico resulta deficiente, y que la actuación de los humanos supera a las fórmulas en tareas de selección y predicción (Holt, 1986). Sin embargo, las conclusiones de Holt han sido duramente cuestionadas debido a las notables dificultades metodológicas y a la falta de confirmación empírica de sus trabajos (Kleinmuntz, 1990).

De otro lado, los resultados de Meehl han sido replicados en numerosos estudios posteriores (por poner algunos ejemplos significativos citaremos autores como: Dawes, Faust y Meehl, 1989; Einhorn, Kleinmuntz y Kleinmuntz, 1979; Kleinmuntz, 1990; Marchese, 1992; Meehl, 1986). Por ejemplo, Goldberg, uno de los defensores de la predicción actuarial más referenciados en el debate fórmula-intuición, inició una serie de trabajos en los que abogaba por la superioridad de la fórmula frente a la intuición (Goldberg, 1965). En sus investigaciones, el autor concluye que para predecir un diagnóstico a partir de las respuestas dadas por los sujetos al MMPI (Minnesota Multiphasic Personality Inventory), resulta más

eficiente una simple combinación lineal de las escalas que la "intuición" de diferentes clínicos "expertos".

Tras establecer la superioridad de las fórmulas sobre la combinación intuitiva de los datos, se inicia un ingente estudio en relación al juicio clínico con dos líneas de investigación fundamentales bien diferenciadas (Goldberg, Faust, Kleinmuntz y Dawes, 1991). De un lado, un gran número de investigadores se interesan por la forma en que razonan los clínicos (especialmente los que se consideran expertos), cómo toman decisiones, qué sesgos y errores cometen en los procesos de razonamiento, qué posibles estrategias se podrían utilizar para corregir estos defectos, y en qué se diferencia la forma de actuar de los clínicos expertos frente a los novatos. A nivel general subrayamos que a medida que se dispone de más datos en esta línea de trabajo, los resultados alertan sobre la existencia de numerosos errores en la forma de proceder de los humanos cuando efectúan valoraciones clínicas, algunos de los cuales parecen realmente difíciles de resolver (Ellis, Robbins, Schult, Ladany y Banker, 1990; Goldstein, 1990; Hawkins y Hastie, 1990; Joseph y Patel, 1990; Koehler, 1991, 1993, 1994; Miller y Karoni, 1996; Smith y Kida, 1991).

Asimismo, y en parte como consecuencia de los resultados obtenidos en la línea de investigación anterior, aparece un área de trabajo cuyo objetivo prioritario es la de facilitar la aplicación de medios técnicos para ayudar y/o sustituir al clínico en muchas de sus tareas de toma de decisiones (First, 1994; Kleinmuntz, 1991). Bajo esta perspectiva, podemos diferenciar a su vez tres líneas de trabajo diferentes: (1) la adopción de algoritmos simples ("paquetes de instrucciones" que permiten resolver un determinado problema a partir de definiciones secuenciales); (2) la incorporación de fórmulas actuariales (fundamentalmente modelos lineales e inferencia bayesiana); y (3) y la adopción de las recientes innovaciones procedentes del ámbito de la inteligencia artificial, que aplicadas a diferentes tareas clínicas complejas han facilitado el surgimiento de la simulación mediante ordenador (fundamentalmente, los sistemas expertos y las redes neuronales). En este marco de trabajo general, los ordenadores han constituido el medio técnico básico para implementar estas nuevas estrategias inferenciales en la toma de decisiones clínicas.

Es importante matizar que cada una de las tres líneas de trabajo enunciadas en el párrafo anterior presenta sus ventajas y también sus limitaciones. Tomemos, como ejemplo, los primeros programas para ayudar en tareas de decisión. Estos sistemas se caracterizaban por incorporar algoritmos que previamente habían sido definidos por clínicos. Estas bases inferenciales pronto denotaron importantes limitaciones: posible falta de consistencia (el hecho de estar basadas en el juicio de los clínicos les puede restar confianza), falta de flexibilidad (una vez definidos, operan de idéntica forma para todos los casos; además, la incorporación o exclusión de elementos en el algoritmo resulta costosa, ya que requiere volver a validarlo) y la imposibilidad de aplicarlos en el caso de incertidumbre (por ejemplo, por la falta de información). De otro lado, el uso de fórmulas actuariales y la simulación mediante ordenador también conlleva sus propias dificultades, algunas de las cuales serán revisadas más ampliamente en los próximos apartados cuando se presenten dichos sistemas.

En los siguientes apartados abordamos brevemente las líneas de trabajo, teóricas y técnicas, aparecidas y consolidadas tras la década de los 50 en torno al juicio clínico, en gran medida como consecuencia de la publicación del texto de Meehl (1954) y del consiguiente debate originado.

2.2.1. Principales acercamientos teóricos al estudio del juicio clínico surgidos tras la confrontación clínico-actuarial

En las últimas décadas se ha iniciado un importante acercamiento al estudio del juicio clínico en general y del razonamiento diagnóstico en particular. En concreto, se ha considerado que existen tres grandes marcos de trabajo que sirven como punto de referencia para dicho estudio (Klein, Orasanu, Calderwood y Zsombok, 1993): la formación de los juicios clínicos, la toma de decisiones y la solución de problemas.

Previamente al desarrollo formal de este apartado, quisiéramos detenernos para señalar algunos aspectos de la exposición que tal vez requieren de una breve aclaración. Creemos que el interés interdisciplinario que despiertan las líneas de investigación que van apareciendo a medida que avanzamos en el desarrollo de este trabajo plantea dificultades formales a la hora de concretar su estudio. Esta situación se advierte especialmente cuando nos proponemos delimitar el área de trabajo concerniente al razonamiento diagnóstico. En este caso, las mayores dificultades se plantean cuando se intentan delimitar y abordar las diferentes orientaciones teóricas desde las que se ha estudiado el juicio clínico y la toma de decisiones.

Debemos matizar que nuestro objetivo a este nivel es meramente introductorio, en la medida en que sólo pretendemos establecer qué líneas generales han guiado el estudio del razonamiento diagnóstico en psicopatología. Sin embargo, este interés por delimitar aspectos a nivel general obliga al uso de términos que tal vez pueden resultar excesivamente amplios y/o de uso controvertido. Creemos también que una discusión más detallada de estos conceptos implicaría la exposición de contenidos que ciertamente nos alejarían de los objetivos generales de nuestro trabajo. Por lo tanto, los siguientes apartados no tienen como finalidad presentar una revisión detallada del modelo paramorfo del juicio iniciado por Hammond (1955) y Hoffman (1960), ni de la teoría de la decisión desde la óptica de la psicología, ni tampoco expondremos extensamente el trazado del proceso y la teoría de la solución de problemas como referentes del juicio clínico. Los siguientes apartados sólo destacan qué marcos teóricos han orientado el estudio del juicio clínico en general y del razonamiento diagnóstico en particular. La razón de esta revisión teórica obedece a que estos referentes conceptuales han servido de guía para el desarrollo y la validación de los sistemas de ayuda clínica (fundamentalmente informáticos) desarrollados durante las últimas décadas, y también han proporcionado claves sobre qué procedimientos utilizar a la hora de entrenar a los clínicos a mejorar sus propios juicios (Kleinmuntz, 1990; Shanteau y Stewart, 1992).

2.2.1.1. Aproximación paramorfa a la formación de juicios diagnósticos

La investigación sobre la formación de juicios diagnósticos constituye un anclaje más dentro de la línea de investigación iniciada por Hammond (1955) y Hoffman (1960), conocida como modelo paramorfo del estudio del juicio clínico.

Los modelos matemáticos lineales suelen ser los procedimientos metodológicos básicos a los que recurre habitualmente este enfoque a la hora de representar el juicio clínico. Es importante matizar que en este caso el objetivo principal de utilizar la formulación matemática no consiste en demostrar la superioridad de los modelos matemáticos sobre los humanos, sino en mostrar la capacidad prescriptiva y predictiva de dichos modelos en relación a las decisiones clínicas (Godoy, 1996).

El principal marco teórico de trabajo al que suele recurrir este enfoque, especialmente en los trabajos de Hammond, es el modelo de la lente y la psicología probabilística de Egon Brunswik (1955). Este modelo tiene como objetivo prioritario representar cómo se perciben y aprenden las características del medio, para lo cual se conceptualiza la relación entre los datos, la relación entre las diferentes dimensiones de la información, la validez del juicio y/o de la

inferencia, y el grado de conocimiento del entorno que posee un sujeto. Bajo este marco de referencia, tanto el medio como la relación sujeto-ambiente se conciben en términos probabilísticos, ya que se postula que los sujetos siempre se encuentran en situaciones de incertidumbre. La ecuación que representa la validez de un juicio clínico es (Hursch, Hammond y Hursch, 1964):

$$V = X_e X_R G + R[(1-X_e^2)(1-X_R^2)]^{1/2}$$

donde X_e representa el sistema del entorno, X_R el sistema de respuesta del sujeto, G el grado de relación que guardan ambos sistemas y R el coeficiente de correlación entre las variables no lineales de los dos sistemas.

Entre los resultados más significativos obtenidos con el modelo en relación a las estrategias de combinación de la información que siguen los sujetos destaca (Brehmer y Joyce, 1988): (1) que los sujetos aprenden (a través del entrenamiento) a utilizar información probabilística, esto es, información cuya validez no es perfecta; (2) que en cualquier medio (probabilístico o determinado) las funciones simples se aprenden antes que las complejas, y la relación directa (positiva) antes que la inversa (negativa); (3) que los sujetos aprenden de forma apropiada indicios múltiples; (4) que el aprendizaje de relaciones no lineales, aún cuando resulta factible, es más lento que el aprendizaje de funciones lineales; (5) que los sujetos pueden aprender, aunque con cierta dificultad, la importancia relativa de los indicios; (6) que es especialmente difícil para los sujetos aprender qué reglas deben utilizar para relacionar de forma adecuada los indicios con el criterio; (7) que la validez de los juicios se relaciona más con la validez de los indicios que con el volumen de información previa disponible; (8) que los efectos de "feedback" que propone el modelo de la lente (comparar el coeficiente subjetivo de utilización de la información con la validez ecológica de los indicios) son más efectivos que decir simplemente al sujeto si ha acertado o no con su juicio; y (9) que los clínicos no superan (ni podrán hacerlo en un futuro) los resultados de la ecuación de regresión, especialmente porque el modelo matemático explica toda la variancia sistemática del criterio que se puede explicar con los datos que se utilizan como predictores (Kleinmuntz, 1993).

2.2.1.2. Estudio del juicio clínico desde la óptica de la teoría de la decisión

Aunque no sea nuestro objetivo principal abordar el estudio del juicio clínico desde una perspectiva histórica, creemos obligado iniciar la exposición de este apartado destacando dos referencias básicas relacionadas con la toma de decisiones desde el punto de mira de la psicología: (1) la aportación de Von Neumann y Morgenstern (1947), quienes son considerados como los precursores del enfoque contemporáneo de la teoría de la decisión; y (2) la de Edwards (1954), que se considera el precursor del estudio de la toma de decisiones en nuestra disciplina.

La teoría de la decisión constituye un área de trabajo característicamente interdisciplinaria, fuertemente influida por la economía y la matemática, cuyo objetivo prioritario consiste en el estudio de la toma de decisiones (Maciá, Barbero, Pérez-Llantada y Vila, 1990). Dentro de esta teoría se puede diferenciar entre un objetivo básicamente psicológico (la descripción de cómo toman sus decisiones las personas) y una faceta normativa o prescriptiva (que se interesa por aprehender los procesos que conducen a las decisiones más óptimas). Precisamente, las interrelaciones que se aprecian al aplicar estos dos enfoques (descriptivo y normativo) en la práctica, limitan su capacidad para definir qué ámbitos de aplicación y qué modelos se incluyen dentro de la teoría (Arce, 1989).

A pesar de las dificultades anteriores, es habitual que los trabajos que se presentan en el ámbito de la teoría de la decisión hagan referencia a los dos enfoques básicos antes

mencionados para estructurar sus contenidos (Martínez Arias, 1991; Payne, Bettman y Johnson, 1992). En este sentido, también es frecuente diferenciar entre dos líneas básicas de trabajo: el análisis decisional y la teoría conductual de la decisión. Esta es la postura que hemos adoptado en este trabajo.

El enfoque del análisis decisional

Se ha considerado que el enfoque del análisis decisional puede ser adoptado con diferentes finalidades, desde la descriptiva a la normativa o prescriptiva. En este sentido, el punto común del estudio del razonamiento clínico desde este enfoque parte de la delimitación de cuatro elementos básicos que intervienen en el proceso de elaboración de los juicios (Mellers, Schwartz y Cooke, 1998): (1) la identificación de las posibles opciones y de las consecuencias de cada opción; (2) la evaluación de la deseabilidad de cada consecuencia (valor); (3) la evaluación de la probabilidad de cada consecuencia si cada una de las opciones fuera adoptada (expectativa); y (4) la combinación de estos elementos de acuerdo a una regla de decisión defendible.

Diferentes trabajos han constatado la utilidad del enfoque decisional como marco de trabajo para el estudio del juicio clínico (Loewenstein y Furstenberg, 1991; Martínez Arias, 1991), ya que lo consideran facilitador del análisis empírico en esta línea de aplicación, de la identificación de las variables que influyen sobre el comportamiento y de la comparación del comportamiento en diferentes dominios y en diferentes situaciones (cuando existen varias alternativas con diferentes probabilidades y preferencias, cuando se valoran situaciones de decisión sobre el uso óptimo de recursos escasos y cuando las situaciones requieren valorar la utilidad de recabar más y nuevos datos en situaciones de diagnóstico). En cada situación, el modelo establece cómo proceder para elegir la mejor alternativa de la forma más racional y coherente posible. Por ejemplo, en el caso de valorar la utilidad de recabar más datos, se toma en consideración tanto la calidad e influencia que puede tener en el juicio diagnóstico la información a recoger (beneficios o pros), como los posibles riesgos y el coste del propio hecho de recoger la información (el coste o contras).

La fundamentación teórica más usual del análisis decisional se encuentra en la teoría de la utilidad esperada, que parte del trabajo de autores como Von Neuman y Morgenstern (1947), y también Savage (1954). Posteriormente, esta línea de investigación se afianza en psicología con los trabajos de autores como Fischhoff, Goitein y Shapira (1982, 1983), quienes la relacionan con los trabajos de Edwards (1954) para gestar modelos de expectativa-valor o de expectativa-valencia.

En cuanto a los resultados obtenidos en la investigación sobre juicio clínico, se ha observado que modelos lineales simples, como los que se establecen desde el seno de la teoría de la utilidad subjetiva esperada, predicen adecuadamente la respuesta de los sujetos, aunque éstos manifiesten que están utilizando otro tipo de reglas (Fischhoff et al., 1982).

La teoría de la utilidad esperada también ha sido objeto de numerosas críticas en su aplicación general al ámbito de estudio del juicio clínico. De un lado, por sus limitaciones e inconsistencias como teoría descriptiva y prescriptiva, ya que se considera que los sujetos están lejos de aplicar lo que establece la teoría cuando toman decisiones en su vida cotidiana (Beach y Lipshitz, 1993; Orasanu y Connolly, 1993). De otro lado, porque garantizar que las decisiones son "lógicamente consistentes" (objetivo final del modelo) no asegura que éstas sean las decisiones mejores que se pueden tomar frente a un problema concreto (Klein, Orasanu, Calderwodd y Zsombok, 1993).

El enfoque de la teoría conductual de la decisión

En gran medida como consecuencia de las limitaciones demostradas por los abordajes basados en modelos normativos de utilidad esperada, han aparecido otros modelos referenciales interesantes en los trabajos sobre elección y toma de decisiones, por ejemplo la teoría conductual de la decisión. Este enfoque delimita una línea de investigación fructífera cuyos inicios se relacionan con la revisión de Edwards (1954), y cuyos resultados denotan su importancia a la hora de aprehender las características del proceso de decisión humano (Martínez Arias, 1991).

Dentro de la teoría conductual de la decisión destaca el concepto cognitivo de "racionalidad limitada", introducido por Simon (1955) para denotar que el hombre no siempre actúa de acuerdo a lo que se establece en los modelos prescriptivos o normativos, sino que se caracteriza por poseer una serie de limitaciones cognitivas que le obligan a construir modelos simplificados del mundo para poder actuar de forma más o menos óptima. De este modo, el autor pretende distinguir claramente la actuación cognitiva humana de las estrategias de solución de problemas consideradas altamente competentes que utilizan otros sistemas lógico-formales (León, 1994).

La teoría conductual de la decisión toma la teoría de la utilidad esperada como un estándar de referencia con el que comparar la actuación de los clínicos, aunque a diferencia de ésta, la investigación sobre decisión conductual estudia la adecuación descriptiva de las teorías normativas (Payne et al., 1992).

Las áreas más investigadas bajo este enfoque han sido (Chase, Crow y Lamond, 1996): (1) los procesos de creación de hipótesis entre las que se debe escoger; (2) la forma en que los sujetos utilizan las probabilidades y las utilidades (propuestas como estándar racional de actuación) para derivar un valor final de la utilidad esperada de cada alternativa; (3) el estudio de las razones por las cuales se producen desajustes entre lo que la teoría de la utilidad esperada (u otros estándares, como el teorema de Bayes) prescribe y la actuación final de los decisores; (4) el estudio de los medios por los cuales se puede ayudar a un sujeto en tareas de toma de decisiones para que escoja la opción más óptima a pesar de sus limitaciones (campo de aplicación del Análisis de Decisiones; León, 1994); y (5) el desarrollo de modelos alternativos a la teoría de la utilidad esperada, como la teoría del coste beneficio (Beach y Mitchell, 1978), la teoría prospectiva (Tversky y Kahneman, 1992) o los sistemas de producción adaptativos (Beach, 1993). En líneas generales, las conclusiones derivadas de numerosos trabajos inscritos en estas áreas de investigación han permitido afianzar la importancia del enfoque de la teoría conductual de la decisión a la hora de comprender las características del proceso de decisión humano. Sin embargo, puesto que gran parte de los trabajos se han llevado a cabo en laboratorio, también se ha despertado cierta suspicacia en relación a la validez externa de numerosas investigaciones.

En cuanto a los resultados obtenidos en los diferentes trabajos llevados a cabo en esta línea, en su mayoría no son demasiado alentadores (Balakrishnan y Ratcliff, 1996). Tras varias décadas de investigación utilizando la teoría de la utilidad esperada, se ha reafirmado la inadecuación del modelo como enfoque prescriptivo del proceso: se ha destacado que la forma en que los sujetos toman sus decisiones difiere sistemáticamente de los supuestos de estos modelos, tal vez por sus limitaciones cognitivas (Payne et al., 1992). A nivel normativo, otro conjunto de trabajos intentan desvelar la adecuación del modelo y las causas de las posibles discrepancias (Keren, 1992; Lopes, 1991).

2.2.1.3. Estudio del juicio clínico desde la óptica de la solución de problemas

A pesar de que los trabajos en torno al trazado del proceso del razonamiento clínico presentan un carácter claramente ateorico, en la actualidad existe un cierto consenso en situar gran parte de las directrices que han guiado esta línea de investigación dentro de la teoría de solución de problemas de Newell y Simon (1972). Esta teoría, también conocida como teoría del procesamiento de la información, constituye un enfoque que, participando de la formulación del trazado del proceso del razonamiento clínico, pretende explicar un hecho de carácter general: los sujetos pueden resolver problemas complejos en un tiempo relativamente breve, y esto a pesar de que el procesamiento de información que caracteriza el sistema cognitivo humano presenta evidentes limitaciones.

En relación al ámbito de estudio del juicio clínico, este esquema general se ha aplicado al proceso de solución de problemas diagnósticos tal como proponen Elstein, Shulman y Sprafka (1990). Estos autores han intentado reflejar cómo se produce el razonamiento clínico (en lugar de establecer por qué procedimientos podría representarse o sustituirse): identificando los procesos cognitivos subyacentes al razonamiento de los clínicos expertos (estableciendo en cada caso si tales procesos dependen de características específicas de la tarea y/o de la situación), generando una teoría psicológica que permita comprender tales características de razonamiento y que se relacione con los conocimientos actuales sobre procesamiento de la información, y desarrollando procedimientos para enseñar a solucionar problemas clínicos. Los autores coinciden en que las etapas principales que destacan en el proceso de resolución de problemas diagnósticos clínicos son: (1) adquisición de la información (a través de la búsqueda y recogida selectiva de datos); (2) generación de hipótesis diagnósticas alternativas; (3) re-interpretación de los datos en función de las hipótesis diagnósticas generadas; (4) comprobación de la validez de las hipótesis generadas; y (5) selección de la hipótesis correcta que se debe emitir como diagnóstico del sujeto.

Existen diferentes estrategias de solución de problemas (Becoña, 1993; Becoña y Gutiérrez-Moyano, 1987), algunas de las cuales han demostrado ser útiles cuando se aplican a diferentes sistemas de ayuda clínica. Entre las más habituales destacan dos (Godoy, 1996; Nezu y Nezu, 1989): (1) la aplicación sistemática sobre el estado inicial y los intermedios de todos los operadores posibles (hasta encontrar el estado que constituye la solución o hasta que se agotan todas las ramas existentes del árbol de decisión y el problema se considera "no solucionable"); y (2) la aplicación de reglas heurísticas (se adoptan si las estrategias anteriores son poco operativas, por ejemplo porque los problemas presentan un cierto nivel de dificultad). De hecho, el empleo de heurísticos en este enfoque es del todo imprescindible si se desean promover acciones competentes. Dos de las reglas heurísticas habitualmente utilizadas son el análisis medios-fin y el encadenamiento hacia atrás.

Finalmente, para que el sistema pueda solucionar de forma competente un problema, debe disponer de dos tipos de conocimiento: declarativo (teórico) y procedimental (práctico o técnico). El primero se compone de las definiciones y descripciones disponibles de conceptos (básicamente de teorías y hechos empíricos), y de sus relaciones. El segundo señala cómo actuar ante una determinada situación (por ejemplo técnicas de evaluación o propuestas terapéuticas), y es habitual que se exprese en forma de reglas de producción o reglas de condición-acción ("si ... entonces ...").

A nivel funcional, se ha demostrado que la práctica en la solución de un determinado tipo de problema produce: rapidez y automatización en el proceso de solución de dicho tipo de problemas, y la compilación de reglas (su unión en unidades más complejas). Esto hace que la acción cada vez sea más independiente del conocimiento declarativo y más dependiente de la situación o del componente procedimental. Este hecho se ha tomado como una justificación de por qué suele resultar difícil determinar cuáles son las actuaciones (óptimas e inconsistentes) de los sujetos que llevan a cabo tareas de decisión clínica.

2.2.1.4. Líneas de investigación futuras y actuales del estudio del juicio clínico

Los acercamientos anteriores al estudio del juicio clínico y de la toma de decisiones han favorecido el surgimiento de diferentes líneas de investigación. A modo de conclusión, destacaremos dos de las más relevantes.

Por un lado, están apareciendo en la literatura diferentes trabajos que tratan sobre la generación de técnicas específicas para entrenar a los sujetos que trabajan en entornos clínicos, en algunos casos mostrándoles cómo usar sistemáticamente algunos heurísticos que mejoran la ejecución en tareas de solución de problemas de razonamiento diagnóstico (Elstein et al., 1990). En concreto, los entrenamientos más habituales han ido dirigidos a aquellos aspectos relacionados con: (1) la recogida de la información (recogida selectiva de datos necesarios, fiables y válidos); (2) la generación de hipótesis (búsqueda activa de hipótesis de trabajo explicativas de un problema, evitando el mantenimiento de una sola hipótesis en el proceso de evaluación y diferenciando en cada momento qué implica la probabilidad de una hipótesis y el riesgo de no tratarlo); y (3) la combinación de la información y la comprobación de hipótesis (contemplar la probabilidad, la utilidad y los beneficios-riesgos de cada hipótesis de trabajo, evitar estrategias de confirmación de hipótesis y agrupar diagnósticos y problemas que puedan abordarse con un mismo programa terapéutico). Asimismo, algunos programas concretos se proponen intentar que los sujetos mejoren su ejecución tomando como modelo los acercamientos actuariales de la regresión y el teorema de Bayes; en este caso lo que se hace es favorecer que los humanos sigan las directrices que prescriben dichos modelos matemáticos. Finalmente, los autores que abordan esta línea consideran fundamental implementar en estos programas de entrenamiento estrategias que incrementen y mejoren el "feedback" que recibe el clínico, que aumenten la motivación de los profesionales para alcanzar resultados de calidad y que informen de que el tipo de tarea que realizan es probabilística.

De otro lado, desde el acercamiento de trazado del proceso y de la solución de problemas, también se han derivado herramientas procedimentales cuyo objetivo consiste en la mejora de la ejecución de los sujetos. En esta línea, ha aparecido una ingente investigación sobre la actuación de los "expertos" que ha ido más allá del ámbito de la psicología para alcanzar el ámbito de la inteligencia artificial, desde donde se están desarrollando diversos tipos de sistemas informáticos de actuación muy prometedora dentro del ámbito de la toma de decisiones clínicas (Hammond et al., 1987; Kleinmuntz, 1990, 1991; Lasala, 1994): los sistemas expertos y las redes neuronales. Más adelante revisaremos estas contribuciones en el plano concreto del diagnóstico psicopatológico.

2.2.2. La adopción de fórmulas: una aplicación de los modelos normativos para simular el juicio clínico en psicodiagnóstico

Tras la publicación del trabajo de Meehl (1954) se han utilizado diferentes procedimientos estadísticos para ayudar al clínico, entre los que destacan: la ecuación de regresión, el análisis de la variancia y el teorema de Bayes. En general, los modelos lineales, y en especial la ecuación de regresión, se han mostrado superiores y han sido los más utilizados, tanto para clasificar y predecir criterios externos como para predecir las respuestas del propio clínico.

La explicación fundamental que se ha dado a por qué funcionan bien los modelos matemáticos, y sobre todo la ecuación de regresión, se debe a que estas fórmulas recogen toda la variancia sistemática posible (Kleinmuntz, 1993), y no se ven afectadas por variables motivacionales, atencionales, perceptuales o socioambientales. Esto es, todo clínico (por

experto que sea) tiene una actuación variable altamente dependiente del contexto; los modelos matemáticos, sin embargo, tienen un comportamiento invariante del plano contextual.

2.2.2.1. *La ecuación de regresión*

La ecuación de regresión que más se ha utilizado para simular los juicios de los clínicos es la propuesta por Hoffman (1960), que representa la contribución relativa de cada predictor (indicio de información disponible) sobre la respuesta de los sujetos (juicios). La expresión matemática del peso relativo (RW_{ks}) de cada ítem de información es: $RW_{ks} = (b_{ks} \times r_{ks}) / R^2$. En esta fórmula, k denota un indicio determinado, b_{ks} representa el coeficiente b de regresión, r_{ks} la correlación entre la característica k y la respuesta del sujeto, y R^2 representa el cuadrado de la correlación entre las respuestas del sujeto y las respuestas que predice el modelo.

Godoy (1996), en una excelente revisión, señala cuáles son las propiedades y las ventajas fundamentales del uso de la ecuación de regresión lineal en el ámbito de la toma de decisiones clínicas. Entre las prestaciones más ventajosas, el autor destaca: (1) que las ecuaciones de regresión que simulan la estrategia de formar juicios de los profesionales reproducen dichos juicios con gran precisión; (2) que las ecuaciones permiten reproducir los juicios de los clínicos y predecir juicios futuros ante nuevos datos; (3) que las ecuaciones representan el juicio de un clínico particular, y también el juicio combinado y conjunto de varios clínicos; (4) que la ecuación generada sobre los juicios de un clínico se puede utilizar para predecir el juicio de otros clínicos (utilizando ecuaciones de regresión impropias); y (5) que el juicio de los clínicos puede ser representado aunque se usen pesos de regresión generados al azar o cuando se eliminan los pesos¹ (asignando 0 a la información irrelevante y 1 a la relevante).

Asimismo, es importante destacar que no se ha conseguido establecer una fuerte asociación entre el valor subjetivo que el clínico asigna a los distintos datos y el correspondiente peso que obtienen en la ecuación de regresión. Esto supone que, o bien la ecuación no representa los procesos cognitivos del clínico, o bien que los profesionales no son capaces de verbalizar (o desconocen) los procesos psicológicos que utilizan al formar juicios. De hecho, cuando se le pide a los clínicos que ajusten el valor subjetivo que asignan a los datos al correspondiente peso en la ecuación de regresión, se mejora el juicio final.

En relación al párrafo anterior, también es preciso señalar que algunos autores han postulado que los clínicos utilizan estrategias diferentes de las lineales al formar los juicios, por ejemplo curvilíneas, configuracionales (no monótonas) o interactivas. Esto implicaría la búsqueda de modelos matemáticos diferentes a la ecuación de regresión si se desea representar el componente procesual de la combinación de la información y la toma de decisiones diagnósticas. Igualmente, en aquellos casos en los cuales la relación entre los predictores y la respuesta (o criterio) no fuera monótona o estable, la ecuación de regresión sería inadecuada, y no podría aventajar en modo alguno la actuación del clínico. Sin embargo, los estudios realizados al respecto indican que, en general, para establecer los juicios que proponen los clínicos no es necesario recurrir a estrategias complejas de combinación de los datos, ni siquiera en los casos en los que dicho análisis fuera posible (Godoy, 1993; Godoy y Silva, 1992). Esto tiene dos explicaciones posibles: que las situaciones en las que se utilizan otras estrategias no lineales suelen ser infrecuentes en la práctica, o también que aunque existan quedan bastante bien representadas a través de modelos más sencillos.

Finalmente, recordar que la ecuación de regresión sólo es adecuada cuando la respuesta (criterio) es cuantitativa. Pero en aquellos casos en los que sólo es posible dar juicios

¹ Esta es la causa de la "intercambiabilidad" de las ecuaciones de regresión para predecir el juicio de otros clínicos: sólo se requiere que los diferentes sujetos consideren por igual qué información es irrelevante-irrelevante y el sentido predictivo de los datos (signo positivo-negativo).

diagnósticos de naturaleza cualitativa, este modelo no es adecuado, y es preciso recurrir a otros procedimientos, como la estadística bayesiana o la simulación con ordenador empleando árboles lógicos.

2.2.2.2. *El análisis de la variancia*

El análisis de la variancia constituye otra aproximación lineal al estudio del juicio clínico. Sin embargo, en principio, este enfoque parece no aportar poder predictivo respecto al uso de la ecuación de regresión. Lo que sí ha demostrado esta técnica estadística es capacidad para detectar y manejar modelos no lineales (curvilíneos o configuracionales), lo cual ha sugerido que podría servir para añadir verosimilitud a la simulación del componente procesual o estratégico del juicio clínico. En efecto, bajo la óptica del análisis de la variancia, los efectos principales de los diferentes indicios se interpretan como "el peso relativo" que los clínicos conceden a dichos indicios, manteniendo constantes el resto de los datos. Cuando se dispone de la suficiente información, se pueden construir modelos que evalúen tendencias, por ejemplo lineal, cuadrática o cúbica. Además, las interacciones entre los factores se suelen interpretar como patrones de indicios, es decir, como un uso configuracional de los factores (en las cuales el valor dado a una variable depende del valor que presenta otra).

En cualquier caso, en la práctica este modelo ha sido poco utilizado en el estudio del juicio clínico. La razón fundamental radica en que los diferentes trabajos han constatado que las interacciones entre los factores explican poca variancia. Existe, sin embargo, una excepción que ha alcanzado cierta resonancia en el estudio sobre formación de juicios: el trabajo de Anderson y colaboradores (Anderson, 1969, 1970) en torno a la teoría de la integración de la información. Esta teoría parte del supuesto general que el comportamiento humano está multicausado, de modo que la formación del juicio requiere integrar un volumen considerable de información entrante antes de ser emitido. Bajo este prisma, dicha integración está dirigida por reglas algebraicas.

El modelo anterior se ha utilizado ampliamente en el estudio sobre toma de decisiones y sobre formación de impresiones. Los objetivos de este enfoque no se limitan a la predicción de los juicios que emiten los sujetos, sino que pretende descubrir las estructuras cognitivas de procesamiento y combinación de la información. Para ello se utilizan reglas aritméticas, ya que se supone que los procesos cognitivos de integración de la información también las siguen.

Hemos indicado que la teoría de la integración de la información de Anderson ha escogido el análisis de la variancia como herramienta estadística a la hora de representar el componente procesual de la toma de decisiones. De acuerdo a este enfoque, cada ítem de información se representa por dos características fundamentales: su valor en una escala subjetiva (s) y su peso o importancia (w). El modelo inicial, que posteriormente se ha modificado para poder ser aplicado a diversos enfoques teóricos del procesamiento de la información, establece que el juicio u opinión de un juez es una función del valor-peso (C) de la opinión previa o inicial (w_0s_0) y el valor escalar e importancia de la nueva información obtenida ($w_1s_1; \dots; w_ks_k$): $R=C+\sum_1(w_ks_k)$.

2.2.2.3. *El enfoque de Bayes*

Muchos trabajos coinciden en que el teorema de Bayes no resulta un fiel reflejo de los procesos cognitivos de pensamiento que utilizan los seres humanos en su quehacer diario (Doherty y Mynatt, 1990; Fear y Healy, 1997; Smith, Penrod, Otto y Park, 1996). Sin embargo, los modelos bayesianos se han utilizado ampliamente, desde el punto de vista prescriptivo, para ilustrar cómo se debería combinar la información de carácter probabilístico, especialmente en el estudio del juicio y de la toma de decisiones (Hekerman, Mamdani, Wellman, 1995).

El modelo bayesiano opera de forma distinta a la del modelo de regresión. Básicamente, el modelo de regresión trabaja a partir de conjuntos fijos de información y tiene como objetivo predecir el valor de una única respuesta en una escala supuestamente cuantitativa, utilizando para ello reglas de combinación lineales-aditivas. Por otro lado, gran parte de los sistemas que trabajan a través de estadística bayesiana parten de reglas previas fijas, y estudian el impacto de la nueva información disponible sobre dicho conjunto. Uno de los objetivos prioritarios de estos sistemas consiste en disponer de una distribución de probabilidad y escoger el valor con probabilidad más alta de ocurrencia. Asimismo, bajo el enfoque bayesiano, el establecimiento del juicio final sigue un conjunto de reglas semejantes a las de la teoría de la integración de la información y la teoría de la medición conjunta: se trata de combinar sucesivamente los datos previos con los presentes, debidamente sopesados.

En la actualidad se dispone de diferentes programas computerizados de ayuda a la toma de decisiones médicas que utilizan el teorema de Bayes (Elwood, 1993; Ennis, Ohmann, Lorenz, Zacyk y Schoning, 1988; De Dombal, 1988). Así, por ejemplo, De Dombal y sus colaboradores pusieron en marcha un programa en el Reino Unido que todavía hoy persiste (De Dombal, 1988), cuyo objetivo consiste en administrar, corregir e interpretar los datos procedentes de una prueba de evaluación para proporcionar un diagnóstico completo de los sujetos con dolor abdominal agudo. En psicopatología también se están utilizando programas que utilizan la fórmula de Bayes con carácter diagnóstico (Lanctôt y Naranjo, 1994; Mitnitski, Graham, Mogilner y Rockwood, 1997; Prince, 1996). La operativa de estos programas es sencilla, basta con que el usuario proporcione al ordenador la correspondiente matriz de datos con las probabilidades asignadas a cada variable relevante para el diagnóstico. A continuación, el sistema evalúa la probabilidad condicional de los síntomas, signos y pruebas de laboratorio proporcionadas, y calcula el vector con las correspondientes probabilidades de que cada paciente posea una y/o otra enfermedad.

Quienes conocen y utilizan estos sistemas saben que el diagnóstico basado en el teorema de Bayes dispone de importantes implicaciones a nivel clínico y epidemiológico. En este párrafo destacaremos una de las prestaciones que, a nuestro juicio, resulta muy atractiva. El proceso de diagnóstico tradicional y convencional maneja con dificultad los problemas de comorbilidad, e intenta priorizar la existencia de un diagnóstico en cada sujeto. Pero bajo la nueva perspectiva, sin embargo, el profesional dispone del conjunto de posibles diagnósticos representados en un continuo de probabilidades, lo cual parece más acorde con la realidad que maneja el clínico y convierte el diagnóstico de cada paciente en un objeto multidimensional.

A pesar de sus ventajas, y del atractivo que supone su uso en determinados contextos, la aplicación práctica de los sistemas que utilizan el teorema de Bayes cuenta con algunas limitaciones (Klein et al., 1993; Krosnick, Li y Lehman, 1990), e incluso algunos autores cuestionan su validez como procedimiento de diagnóstico. En los siguientes párrafos se destacan las dificultades más sobresalientes del modelo.

Algunas de las limitaciones que tradicionalmente se han relacionado con el uso del teorema de Bayes reside en su aparente complejidad y en los requisitos de cómputo. Sin embargo, y puesto que los modelos matemáticos siempre se usan con la ayuda de ordenadores (los cuales resuelven totalmente estas dificultades técnicas), actualmente se considera que las

mayores limitaciones del modelo derivan del tipo de datos que requiere conocer: la probabilidad previa del evento, la información adicional específica sobre el evento y la probabilidad esperada de acierto (o precisión esperada en la predicción). En este sentido, los detractores del modelo apuntan que los sistemas que incorporan el teorema de Bayes como regla de decisión trabajan, en su mayoría, con estimaciones aproximadas o incorrectas, y que no suelen tener en cuenta la fiabilidad y la diagnosticidad de la información disponible (Doherty y Mynatt, 1990). Lo que es cierto es que en la actualidad los modelos basados en la estadística bayesiana requieren disponer de un volumen de información considerable, información que de otro lado suele ser recabada de forma lenta y sin atender a unas reglas normativas. Como apuntamos en la presentación de este trabajo, en la práctica es difícil disponer de datos fiables sobre las probabilidades previas, y en muchos casos el valor escogido para la fórmula depende de la elección particular que realiza cada clínico ante un conjunto de valores potenciales. Esto implica que los resultados dados por el sistema son dependientes del usuario particular, o lo que es lo mismo, son idiosincrásicos a cada clínico. Es más, el valor de probabilidad único asignado a una hipótesis no indica nada acerca de su precisión; por ejemplo, declarar que $p(x|A)=0.5$ podría significar tanto 0.5 ± 0.4 como 0.5 ± 0.001 .

Por otro lado, es importante tener presente que la aplicación de la fórmula de Bayes, como la de cualquier modelo matemático, requiere que se cumplan unos determinados supuestos. En este caso, la restricción fundamental para aplicar el teorema consiste en que los diferentes descriptores se relacionen estructuralmente (independencia condicional). Sin embargo, puesto que en nuestra disciplina la estructura entre los síntomas suele ser desconocida, se podría considerar que la aplicación del teorema resulta, cuanto menos, cuestionable.

Finalmente, se ha sugerido que la validez de los resultados obtenidos con métodos bayesianos es tan cuestionable como la selección de síntomas (signos) utilizados como descriptores de los correspondientes diagnósticos (Graham et al., 1996). Aunque sería deseable realizar la selección de signos en función de la evaluación de jueces expertos y de la validez de contenido y de constructo demostrada empíricamente, en práctica los descriptores se suelen escoger en función de la disponibilidad de información de la base de datos. Puesto que una selección diferente modificaría las probabilidades finales, los resultados pueden ser cuestionados en cualquier momento. En relación a este punto, también se ha argumentado que en ocasiones resulta difícil modificar un conjunto bayesiano de valores debido a las dependencias entre dichos valores. Dado que las probabilidades de todas las hipótesis probables que revelan la evidencia E suman 1, si se añade una nueva hipótesis se deben volver a recalcular muchos otros valores y modificar el algoritmo computacional.

En síntesis, a pesar de que la estadística bayesiana se ha aplicado de forma muy fructífera al estudio psicológico de los procesos cognitivos humanos, en la actualidad se dispone de pocos trabajos que apoyen la validez de este enfoque como modelo prescriptivo del juicio clínico. Así, por ejemplo, cuando en su día Grove (1985) aplicaba el teorema de Bayes a través de un programa de simulación para establecer el punto de corte de los criterios diagnósticos de las categorías del DSM-III, el autor dijo no encontrar mejoras respecto a las normas establecidas en esa misma clasificación.

Bajo nuestro punto de vista, algunas de estas matizaciones señaladas como limitaciones del modelo bayesiano podrían generalizarse a otras técnicas estadísticas, ya que sirven para señalar ciertos aspectos relacionados al uso de los modelos matemáticos. Para que un modelo estadístico funcione de forma adecuada, es imprescindible facilitarle la información que requiere codificada de forma adecuada. En este sentido, aún cuando se viene afirmando que las fórmulas estadísticas superan la actuación de los clínicos en las tareas de clasificación y predicción, esta mejora se debe al manejo y combinación que dichos modelos realizan de la

información. Lo que resulta del todo imprescindible es que el clínico facilite los datos requeridos con el formato de registro adecuadas.

Por lo demás, creemos que un estudio riguroso de las posibilidades que ofrece la aproximación probabilística (estadística) en psicodiagnóstico resulta muy productivo desde un plano tanto teórico como práctico. Así, por ejemplo, ya hemos destacado que al calcular el vector de probabilidades utilizando el teorema de Bayes se contemplan "todos" los posibles diagnósticos para cada sujeto. Esto podría orientar al clínico para la implementación de terapias muy específicas según las necesidades particulares de cada paciente, sobre todo porque se deja de atender al diagnóstico primario (o más probable) y se tienen en cuenta las posibles comorbilidades. Además, la disponibilidad de este vector facilita la exploración de los casos considerados "anómalos", es decir, aquellos que manifiestan cuadros sintomáticos atípicos que no se contemplan en las clasificaciones descriptivas habituales.

2.3. INTRODUCCIÓN DE PROGRAMAS INFORMÁTICOS CONVENCIONALES COMO HERRAMIENTAS DE AYUDA EN LA EVALUACIÓN PSICOLÓGICA

A pesar de la superioridad empírica demostrada por las estrategias actuariales de combinación de la información, hoy en día clínicos e investigadores se muestran ajenos ante este cúmulo de información (o simplemente la desconocen), y en la práctica continúan estableciendo conclusiones basadas en sus propios juicios y predicciones (Dawes, 1993; Dawes et al., 1989). En este sentido, aunque la polémica suscitada en los años 50 en torno a la predicción clínica/actuarial ha decaído en las últimas décadas, la realidad práctica y las causas que dieron lugar a tales debates parecen no haber cambiado demasiado desde entonces (Kleinmuntz, 1990, 1991, 1993; Marchese, 1992). Esto supone que hoy en día el problema subyacente a la falta de calidad de los juicios clínicos todavía radica en los errores y defectos que se cometen al combinar la información, y no tanto en la baja calidad de los datos que se manejan en la clínica. Consiguientemente, si el profesional se propone mejorar los resultados de la evaluación diagnóstica necesita mejorar tanto los instrumentos y los procedimientos de recogida de datos como los sistemas de gestión y combinación de la información. Con este objetivo, paralelamente a los intentos de adoptar las fórmulas estadísticas, nos encontramos los esfuerzos de numerosos profesionales para adoptar el ordenador como una herramienta más de ayuda en el proceso de evaluación y diagnóstico psicológico.

La disponibilidad de ordenadores ha tenido un enorme impacto en el ámbito de las Ciencias de la Salud, aunque en diferente grado según las áreas de conocimiento. Los ordenadores fueron introducidos inicialmente en medicina, en torno a los años 50, como un instrumento de ayuda a la práctica clínica. A lo largo de estas últimas décadas se ha producido en este área un interés creciente y un notable avance en los sistemas computerizados de evaluación y diagnóstico. En psicología, los ordenadores fueron adoptados con posterioridad, en torno a los años 60, y el impacto fue considerablemente menor. Entre las causas que han contribuido a este desequilibrio destacan dos factores (Adarraga y Zaccagnini, 1992; Binik, Servan-Schreiber, Freiwald y Hall, 1988): (1) el menor estatus social del que todavía goza nuestra disciplina; y (2) la sensación de que los conocimientos psicológicos todavía son vagos y los métodos de diagnóstico aún no están lo suficientemente estructurados para poder ser afrontados satisfactoriamente desde la informática. Afortunadamente, en la actualidad los sistemas informáticos se están extendiendo a ritmos acelerados a prácticamente todos los aspectos de la evaluación y el diagnóstico psicológico, desde la construcción de instrumentos de medida hasta la estandarización, corrección e interpretación de las pruebas.

Tal como veremos en lo que resta de capítulo, se pueden establecer al menos dos etapas bien diferenciadas en relación a la incorporación de aplicaciones informáticas en evaluación psicológica (Watkins y McDermott, 1991). En una primera generación, el ordenador se utiliza

para crear programas "sencillos" o "convencionales", cuyo objetivo no es otro que adaptar las principales pruebas de evaluación disponibles con fines clínicos y/o epidemiológicos. En esta etapa, el ordenador se utiliza principalmente para facilitar la corrección de dichas pruebas y generar los correspondientes informes psicológicos. En una segunda generación, y debido a la influencia de disciplinas como la psicometría, la inteligencia artificial o la psicología cognitiva, aparece un ingente volumen de trabajos que presentan las ventajas de aplicar sistemas más sofisticados para la ayuda a las tareas habituales de toma de decisión diagnóstica: pruebas adaptativas computerizadas, sistemas expertos (aplicaciones informáticas que modelizan el comportamiento de un sujeto "experto" que realiza una tarea intelectual en un área concreta, por ejemplo la toma de decisiones clínicas y/o diagnósticas) y redes neuronales.

Evidentemente, el desarrollo de los diferentes programas que se han venido utilizando en evaluación psicológica está directamente vinculado al avance tecnológico en software y hardware (Lockshin y Harrison, 1991). En este sentido, puesto que los primeros ordenadores utilizaban una infraestructura aparatosa y enormemente cara, los sistemas informáticos fueron desarrollados para ser utilizados por unos pocos profesionales en áreas muy concretas de la evaluación psicológica (habitualmente, la corrección de pruebas, el análisis cuantitativo de los resultados y la generación de informes psicológicos y diagnósticos). A medida que los ordenadores constituían herramientas más accesibles a diferentes usuarios y profesionales, y también con el incremento del interés por la orientación actuarial en la interpretación de los resultados obtenidos en las diferentes pruebas, los ordenadores comenzaron a utilizarse de forma más generalizada a lo largo de todo el proceso de evaluación. En la actualidad, se han comenzado a diseñar programas capaces de ayudar en múltiples y complejas tareas clínicas. El rápido desarrollo de la informatización ha permitido que en muy poco tiempo aparezcan nuevas técnicas de medida y manejo de la información en psicología, algunas de las cuales tienen importantes repercusiones en la forma de recogida de los datos: la aplicación de entrevistas estructuradas informatizadas, la nueva generación de técnicas basadas en la Teoría de Respuesta al ítem y los avances en el área de la inteligencia artificial.

Y este es el panorama que disfrutamos en la actualidad. Debido a la disponibilidad de ordenadores portátiles muy potentes y fáciles de utilizar, los sistemas informáticos se comienzan a considerar como elementos indispensables para la administración, corrección e interpretación de las baterías de evaluación psicológica más habituales. Entre los programas de este tipo existentes hasta la actualidad se diferencian aquellos que utilizan reglas actuariales contrastadas empíricamente (como muchas de la empleadas a la hora de interpretar pruebas de personalidad) de aquellos otros que emplean árboles lógicos de decisión derivados de forma intuitiva a partir de los conocimientos clínicos (Frost, 1989; Godoy, 1996; Hand, 1990; Lasala, 1994; Madsen, Reinke, Fehrs y Yolton, 1991).

2.3.1. Fases de introducción del ordenador como herramienta de ayuda en el ámbito de la evaluación psicológica

En este apartado presentamos, a modo de resumen, un breve recorrido histórico sobre el uso programas informáticos "convencionales" en el ámbito de la evaluación psicológica. Las diferentes secciones en las que se organiza la exposición corresponden a las etapas que se derivan de los avances tecnológicos en informática y al grado de disponibilidad de los nuevos sistemas informáticos para los usuarios de la salud mental, tal como se presentan en el trabajo de Honaker y Fowler (1990).

Queremos destacar que en lo que resta de este proyecto nos centraremos en el uso de la tecnología informática en lo que se refiere a entrevistas personales ("Computer-Assisted Personal Interview" o CAPI). Si bien es cierto que entre los primeros usos del ordenador

destaca la entrevista a través del teléfono ("Comptuer-Assisted Telephone Interviewing" o CATI), en nuestro trabajo no abordamos esta aplicación debido a que en la actualidad su uso se ha extendido fundamentalmente al ámbito de la investigación en actividades mercantiles o sociales, y su impacto ha sido mucho menor en el ámbito del psicodiagnóstico. Hemos de puntualizar, sin embargo, que los trabajos publicados que indican haber utilizado esta tecnología aportan resultados satisfactorios (Kobak et al., 1997; Rohde, Lewinsohn y Seeley, 1997).

Asimismo, la exposición de este capítulo se centra en los sistemas desarrollados en Estados Unidos. La razón fundamental estriba en que este país se ha constituido como el principal creador de sistemas de evaluación informatizados, de modo que muchos otros países, entre ellos España, se han limitado a importar dichos productos.

2.3.1.1. Primera etapa: década de los 60

Esta primera etapa se caracteriza por el uso de grandes ordenadores, herramientas muy costosas y enormemente aparatosas que requerían de una gran infraestructura para su instalación y mantenimiento. Si bien se trata de una fase inicial en la que quedó claramente establecida la utilidad de los ordenadores como herramientas para la corrección e interpretación de pruebas psicológicas, ya hemos indicado que el acceso de estas nuevas tecnologías se limitó a una pequeña parte de los psicólogos, de forma que el impacto de los nuevos sistemas informatizados no fue muy notorio.

Es importante señalar que aunque la primera generación de ordenadores surgió a finales de la segunda guerra mundial, los sistemas informáticos no comenzaron a utilizarse de forma habitual en ambientes académicos (tales como universidades) hasta mediados de los años 50. De hecho, fue en la década de los 60 cuando los responsables de la gestión sanitaria, en colaboración con los responsables de empresas de suministro informático, comenzaron a explorar las aplicaciones de la tecnología informática en el ámbito de la salud mental. Esta colaboración estimuló dos grandes aplicaciones: sistemas informáticos para la gestión administrativa y programas piloto como herramientas de ayuda durante el proceso de evaluación psicológica.

Los primeros intentos por informatizar la evaluación psicológica se hicieron con pruebas de personalidad ampliamente utilizadas que resultaban especialmente tediosas de corregir e interpretar, como el MMPI, el 16PF y, posteriormente, el Rorschach (Moreland, 1990a, 1990b). A inicios de la década de los 60 aparece el primer gran programa diseñado para ser usado durante el proceso de evaluación psicológica: el Sistema Mayo, de la Mayo Clinic en Minnesota (Swenson, Rome, Pearson y Brannick, 1965). El Sistema Mayo ofrecía las puntuaciones del MMPI y una breve interpretación de aquellas escalas en las que se observaban puntuaciones elevadas. A pesar de ser enormemente sencillo, el programa permitía evaluar una gran cantidad de sujetos que de otra manera no hubieran gozado de atención psicológica o psiquiátrica debido a las carencias de recursos y personal sanitario.

El trabajo realizado en Minnesota, de carácter meramente práctico, aceleró el desarrollo de nuevos programas de interpretación de pruebas psicológicas. Estos nuevos sistemas diferían en cuanto al estilo y la forma de trabajo, aunque compartían los mismos objetivos. Y a diferencia del Sistema Mayo, que fue construido para interpretar cada escala del MMPI por separado, los nuevos sistemas ofrecían una interpretación más global de las pruebas de evaluación, valorando la combinación de las puntuaciones de cada subescala.

Durante esta década también se desarrollaron otros sistemas informatizados de orientación dinámica para interpretar el MMPI, caracterizados por incluir indicaciones terapéuticas apropiadas a cada caso (Caldwell, 1971; Finney, 1965, 1966). Asimismo, uno de los programas que recibió bastante interés profesional fue el desarrollado en 1963 por Fowler

(Fowler, 1964, 1965, 1969), quien pretendía construir un sistema que simulara el proceso de toma de decisión que realizaría un experto frente a las puntuaciones obtenidas en el MMPI, y también ofrecer un informe similar al que redactaría un clínico (en estilo y contenido).

2.3.1.2. Segunda etapa: década de los 70

Esta fase se caracteriza por la aparición de ordenadores más potentes que los de la fase anterior, mucho más económicos, reducidos en tamaño y, por consiguiente, más fácilmente accesibles para el psicólogo medio. Esto contribuyó a incrementar notoriamente el uso del ordenador en diferentes ámbitos de la salud mental.

Se debe tener presente que durante la etapa anterior los responsables del diseño de los sistemas informáticos no habían trabajado en colaboración con los profesionales de la psicología y la psiquiatría, de tal manera que los diferentes programas disponibles hasta entonces se habían desarrollado sin tener presente qué elementos eran los necesarios para mejorar el proceso de evaluación psicológica. Esto hizo que a inicio de los años 70 muchos clínicos se mostraban aún escépticos respecto al uso de los ordenadores en los servicios de salud mental. Como respuesta al descontento de muchos usuarios comenzaron a aparecer sistemas informáticos mucho más complejos que realizaban otras tareas además de la corrección y la interpretación de las pruebas de evaluación, como por ejemplo la administración directa de los reactivos con el ordenador.

A inicios de los 70, ya se disponía de diversas pruebas para administrar directamente, corregir e interpretar varias pruebas psicológicas (Johnson, Giannetti y Williams, 1976, 1978, 1979; Johnson y Williams, 1975, 1978). Estos programas solían incluir distintas baterías, que eran seleccionadas por los profesionales para ser administradas a los pacientes durante una primera evaluación. Era habitual que cada sujeto contestara el MMPI, un test de inteligencia, un test de memoria, un listado de conductas problema, un test de sintomatología depresiva y un examen estructurado del estado mental que llevaba a cabo un profesional y cuyas respuestas se registraban directamente en una terminal (Johnson y Williams, 1975).

En 1977 se fundó el Psych Systems, la primera compañía que comercializó hardware y software para administrar, corregir e interpretar baterías de evaluación psicológica. El software, conocido como Fasttest, contenía 30 instrumentos de medida, incluidas pruebas de inteligencia, interés vocacional, personalidad, historia médica o abuso de sustancias.

2.3.1.3. Tercera etapa: década de los 80

Esta etapa se caracteriza por el notable incremento en el uso de los ordenadores como herramientas de ayuda en el proceso de evaluación psicológica, así como también por la comercialización de muchos programas informáticos especialmente diseñados con este fin. El motivo principal se encuentra en la notoria reducción de los costes de los productos informáticos, los cuales pasaron a estar controlados por grandes empresas comerciales.

En 1981 CompuPsych presentó uno de sus productos principales, Psychometer, un pequeño ordenador portátil diseñado especialmente para administrar, corregir e interpretar diferentes pruebas psicológicas, entre ellas el MMPI. Pero aunque Psychometer tuvo bastante éxito, muchos clínicos se sentían más atraídos por tecnología informática que reunía en un único aparato pruebas de evaluación psicológica, software de procesamiento de textos, de administración, de gestión de datos, etc. Fue de este modo como otras compañías (incluidas Psych Systems, Psychological Assessment Resources o Precision People) desarrollaron y comercializaron software para los ordenadores portátiles más utilizados, Apple e IBM.

En 1982 aparece Psychological Software Specialists (PSS), una de las primeras compañías que publicó un catálogo con una selección de diferentes tipos de software especialmente diseñados para ordenadores personales y que tenían como objetivo la evaluación psicológica. Durante ese mismo año también se fundaron otras dos compañías que contribuyeron al desarrollo de software de evaluación: Psychologist y Applied Innovations (AI). A partir de ese momento proliferaron las pequeñas compañías cuyo objeto consistía en el desarrollo y la comercialización de software de evaluación psicológica. Una guía desarrollada por la Association for the Measurement and Evaluation in Guidance (1984) contenía 112 productos informáticos con aplicaciones en evaluación psicológica. Y el listado presentado por de Krug (1984) incluía 190 productos con aplicaciones en evaluación y modificación de la conducta. Los productos informáticos de evaluación psicológica estaban cada vez más próximos y eran más utilizados por clínicos e investigadores. A su vez, este hecho atrajo la atención de empresas mayores, como la National Computer Systems (que acabó por adquirir CompyPsych).

Durante la segunda mitad de la década de los 80, numerosas empresas continuaron desarrollando y comercializando pruebas de evaluación especiales para psicología. Así, por ejemplo, en 1987 ya se disponía de más de 300 aplicaciones informáticas catalogadas (Krug, 1987). Muchas de ellas estaban destinadas a la administración (en concreto a la autoadministración), corrección e interpretación de los tests tradicionales en lápiz-papel.

2.3.1.4. Estado actual: los sistemas convencionales ceden protagonismo a los sistemas complejos

De lo expuesto anteriormente se deduce que en sólo unas pocas décadas se ha producido un notable aumento en la aparición de instrumentos de evaluación psicológica. En los últimos años, esta tendencia se ha incrementado, en parte como consecuencia del importante avance en las prestaciones de hardware de los ordenadores. Aunque el desarrollo del software no ha seguido una trayectoria tan espectacular, en esta década también han proliferado programas de notable interés científico para psicólogos y psiquiatras (Conoley, Plake y Kemmerer, 1991; Lewis, 1992).

En la actualidad los ordenadores se utilizan en diferentes áreas fundamentales de la evaluación psicológica: evaluación de la personalidad, evaluación de las facultades cognitivas, desarrollo de entrevistas informatizadas, autorregistro de conducta, registro de datos psicofisiológicos, desarrollo de programas de entrenamiento de los sujetos y programas de ayuda a la toma de decisiones clínicas (Farrell, 1991). Los sistemas informáticos han cumplido diferentes funciones en estos ámbitos, entre las que destacan (Bloom, 1992; Jäger, 1991; Kleinmuntz, 1991): la recogida de información psicológica, el estudio de los posibles factores de riesgo, el estudio del ajuste psicosocial de los individuos y la valoración de los cambios en las conductas y en la sintomatología. Cada vez con mayor frecuencia, están apareciendo trabajos que indican que el ordenador debería formar parte integral de todo el proceso de evaluación psicológica (Honaker y Fowler, 1990).

En los siguientes subapartados resumimos algunas de las prestaciones más innovadoras de la tecnología informática al ámbito de la evaluación psicológica actual. No nos referiremos a las entrevistas estructuradas debido a que más adelante ya se les dedica todo un capítulo al ilustrar el proceso que hemos seguido para adaptar uno de estos protocolos en formato informatizado.

Evaluación de la personalidad, la inteligencia y las funciones neuropsicológicas

Los procedimientos de administración directa e interactiva de las pruebas de personalidad, de los tests de inteligencia y de las baterías de evaluación neuropsicológica han gozado desde hace tiempo de un gran interés por parte de muchos investigadores, y en la actualidad representan el mayor número de las aplicaciones existentes. Anteriormente indicamos que en un principio estos programas de administración directa se desarrollaron para tests objetivos de personalidad, tales como el MMPI, el test de Eysenck o el 16PF, posiblemente debido a que los reactivos eran muy simples en estas pruebas (meros enunciados verbales) y las opciones de respuesta estaban claramente limitadas. Muchos de estos programas presentaban simultáneamente en la pantalla las preguntas y las opciones de respuesta, de modo que el paciente se limitaba a escoger la opción deseada con una pulsación (o a saltar la pregunta si lo desea). Posteriormente también se han realizado notables esfuerzos para administrar algunas pruebas proyectivas, como el test de Rorschach, y en la actualidad existen programas para administrar la mayoría de las pruebas de evaluación de la personalidad.

En el caso de la evaluación de la inteligencia y de los componentes neuropsicológicos, debido a la complejidad de muchas de las pruebas existentes (escalas de Weschler, matrices progresivas de Raven, Peabody, etc.), los programas de administración directa han sido menos numerosos que los de evaluación de la personalidad. De hecho, durante algún tiempo muchos de los programas disponibles en el mercado se han limitado a la administración de las pruebas que miden habilidades verbales (Bradley, Welch y Skilbeck, 1993; Kane y Kay, 1997; Saz, Dia, de la Camara y Carreras, 1996).

En su mayoría, los programas existentes para evaluar la personalidad, las funciones intelectuales y los procesos cognitivos han sido diseñados adaptando los tests tradicionales existentes. En la actualidad, sin embargo, también están apareciendo tests adaptativos para obtener una estimación de los atributos deseados minimizando el tiempo de administración de las pruebas.

Finalmente, la mayoría de los programas diseñados para administrar estas pruebas de evaluación también disponen de procedimientos de corrección e interpretación. En este sentido, existe una gran diversidad de software que permite calcular puntuaciones directas, puntuaciones típicas y realizar diversos análisis psicométricos. Muchos de estos programas incluyen algoritmos para obtener puntuaciones diferentes a las habituales, medidas que por su complejidad muchas veces no se obtenían cuando las pruebas se corregían manualmente. Asimismo, y a pesar de que muchos programas se han desarrollado para corregir un test concreto, los sistemas más recientes corrigen e interpretan simultáneamente el conjunto de resultados obtenidos en varias pruebas.

Desarrollo de programas para el seguimiento y la evaluación de terapias

Además de la recogida, combinación e interpretación de los datos, los ordenadores se han comenzado a utilizar como valiosos instrumentos de colaboración con los programas terapéuticos, por ejemplo para la recogida sistemática de los datos de los pacientes en su entorno habitual o para facilitar el control de conductas.

Para ilustrar este apartado nos serviremos de un ejemplo obtenido en población adulta. Agras, Taylor, Feldman, Losch y Burnett (1990) describen un programa para el tratamiento de la obesidad que incluye el control diario de la ingesta de calorías a través de un microordenador. Cada mañana, los sujetos introducen cuáles son sus objetivos sobre el número de calorías que tomarán y el peso. Posteriormente, durante cada comida registran las calorías que ingieren. Los sujetos pueden disponer en cualquier momento del total de calorías

ingeridas y de gráficos-resumen del total de ingesta diaria y del peso. Además, el ordenador propone "menús" para cumplir con los objetivos propuestos, mensajes disuasorios de ingesta alimentaria y mensajes reforzantes de conductas de control.

El uso de microordenadores como técnica para el registro y control de conductas posee numerosas ventajas (Colby, Gould y Aronson, 1989; Lawrence, 1986; McCullough, Farrell y Longabaugh, 1986; Sampson, 1986). Entre ellas destaca su capacidad para contrastar y facilitar la adhesión de los sujetos a las terapias establecidas. Además, ofrece mayor posibilidad para realizar análisis que contemplen la variable "tiempo" (cuándo aparece un dolor más severo, cuánto duran las crisis de ansiedad, etc.). Este procedimiento también mejora la precisión de los datos, ya que las respuestas inconsistentes que entran los sujetos al ordenador pueden ser detectadas y corregidas de inmediato. De otro lado, cuando se emplean programas de simulación los pacientes pueden examinar los efectos y las consecuencias de las estrategias disponibles para resolver sus problemas.

Finalmente, algunos autores señalan que no todos los pacientes se benefician por igual del uso de ordenadores como complemento y/o director de las terapias psicológicas, y que podría ser inadecuado adoptar esta estrategia en el caso de pacientes orgánicos cerebrales o en sujetos con diagnóstico de psicosis (Colby et al., 1989). Sin embargo, desde hace más de una década se pueden hallar en la literatura un número considerable de ejemplos que demuestran que es posible desarrollar programas especiales para personas que presenten muy diversas dificultades físicas (por ejemplo sensoriales) y/o psicológicas (Dattilo, 1986; Pace, Ivancic, Edwards, Iwata y Page, 1985; Wacker, Berg, Wiggins, Muldoon y Cavanaugh, 1985). Esta posibilidad se hace cada vez más real con las prestaciones que en la actualidad ofrecen las nuevas tecnologías informáticas. En cualquier caso, siempre será responsabilidad del clínico seleccionar aquellos sujetos que pueden beneficiarse de un tipo concreto de intervención.

Recogida de datos de observación

La recogida de datos a partir de la observación directa resulta una labor, cuanto menos, tediosa. Con el desarrollo de ordenadores portátiles esta tarea se ha simplificado notoriamente. En este sentido, registrar este tipo de información implica, simplemente, realizar unas pulsaciones en el teclado, y si es preciso, las anotaciones que el observador considere oportunas (Repp, Harman, Felce, Acker y Karsh, 1989).

Los programas diseñados con este propósito poseen numerosas posibilidades. Por ejemplo, pueden ofrecer resúmenes estadísticos al final de cada sesión sobre la frecuencia y/o duración de cada conducta, así como también de la relación entre las variables registradas. La disponibilidad de microordenadores constituye otra de las prestaciones que en la actualidad resulta más atractiva. Eiler, Nelson, Jensen y Johnson (1989) presentan un ejemplo en el que utilizan un pequeño "scanner" para leer códigos de barras que representan las diferentes conductas a observar. Los autores demuestran que este sistema requiere menos entrenamiento e incrementa la precisión en comparación a los portátiles que se sirven de teclados para la grabación de datos.

Recogida de datos psicofisiológicos

Los ordenadores poseen una larga tradición en el registro de medidas psicofisiológicas. Sin embargo, en su inicio, se trataba de sistemas enormemente caros y con limitaciones evidentes. El desarrollo de ordenadores más sofisticados durante la última década ha facilitado enormemente el registro de estas variables. Los programas existentes actualmente permiten obtener medidas de forma casi constante, y también analizarlas de forma simultánea (García y

Roa, 1995). Pensemos que la evaluación psicofisiológica suele implicar el manejo de un gran número de variables y valores, razón por la cual dichas variables no siempre han gozado de la atención necesaria en los procesos de evaluación psicológica (Vila, 1990). Puesto que el uso de ordenadores facilita el manejo y análisis de datos, es muy probable que en un futuro no lejano estas variables formen parte integral de dicha evaluación (Golu y Bogatu, 1993; McArthur, Schandler y Cohen, 1988).

Entrenamiento de los futuros clínicos

El uso de ordenadores con fines docentes se remonta a la década de los cincuenta, aunque su adopción en el seno de la psicopatología es algo más reciente. La inclusión de programas de entrenamiento en el ámbito de la medicina y la psicología clínica ha adoptado dos ópticas diferentes. Hemos visto que, de un lado, los ordenadores se han utilizado para instruir a los pacientes en el proceso de evaluación y terapia (Sampson, 1986), y también para la instrucción de la población en ciertas conductas de prevención y de riesgo de psicopatología (Bloom, 1992).

De otro lado, los ordenadores se pueden utilizar para el entrenamiento de los futuros clínicos (Schoenfeldt, 1994). Esta última línea ha gozado de un creciente interés durante los últimos años, y cada vez es más habitual que los diseñadores de "software" preparen programas de simulación que permitan a los futuros profesionales aprender y ejercer las habilidades requeridas en evaluación y diagnóstico.

Asimismo, los ordenadores que se utilizan como herramienta de entrenamiento pueden dirigirse a diferentes aspectos fundamentales del conocimiento. Jäger (1994) señala tres áreas fundamentales: la procedimental, la declarativa y los aspectos metacognitivos de conocimiento.

Veamos algunos de los trabajos aparecidos en esta última línea de investigación. Lichtenberg, Hummel y Shaffer (1984) presentaron un modelo de primera entrevista computerizada a la que denominaron CLIENT 1. El programa simulaba un paciente con problemas en diferentes áreas como el trabajo, la familia, las emociones y sus relaciones sociales. En este caso, la misión del futuro clínico consistía en determinar cuál era el área que generaba el mayor número de problemas o el que más preocupaba al supuesto paciente.

Más adelante, Lambert, Hedlund y Vieweg (1990a, 1990b) presentan otro ejemplo de programa de entrenamiento. En este caso, los autores habían programado en un ordenador diferentes tipos de problemas y diferentes condiciones. Durante el entrenamiento, los sujetos van escogiendo las diferentes opciones y estableciendo sus propios juicios. El sistema ofrece feedback inmediato sobre las elecciones de los sujetos, y también un resumen final sobre la estrategia seguida durante la sesión, sus pros y contras.

En la actualidad, se están multiplicando los trabajos que presentan programas informáticos para mejorar las habilidades de los futuros clínicos en tareas de diagnóstico y toma de decisiones, tanto en los ámbitos de la medicina como en la psicología clínica (por poner sólo algunos ejemplos remitimos al lector a los trabajos de: Fincher et al., 1986; Florey, 1988; Garret, Savage y Hendrickson, 1990; Kidd, Cesnik, Connoley y Carson, 1992; Liras, 1990; Tood, 1996; Walsh y Bohn, 1990; Wigton, 1987; B.P. Wolf, 1992). En conjunto, estos trabajos constatan una clara aceptación de las nuevas técnicas de enseñanza por parte de los sujetos, y también evidencian que complementar el entrenamiento de los futuros clínicos con programas de ordenador ofrece un notable número de ventajas. Así, por ejemplo, el aprendizaje con ordenador puede dirigirse de forma concreta a las necesidades específicas de cada sujeto. De este modo, el uso de la informática permite que los futuros clínicos sean agentes activos en el proceso de formación profesional, a la vez que les permite definir, imponer y seguir su propio ritmo de aprendizaje. Siguiendo en esta misma línea, los usuarios

de los programas informáticos de entrenamiento destacan que el ordenador es un instructor paciente, objetivo e imparcial, que ofrece feedback de forma consistente a la ejecución de los sujetos (éstos pueden realizar errores de forma involuntaria o voluntaria, con objeto de evaluar la consecuencia de sus acciones), que modifica de forma relativamente simple las instrucciones en función de las respuestas previas, que integra un gran volumen de información de forma estructurada y flexible, que incrementa notoriamente la motivación de los estudiantes, que es accesible a cualquier hora del día, y que permite conectar entre sí diferentes zonas geográficas (por ejemplo, diferentes universidades, o zonas rurales con zonas urbanas). Además, el entrenamiento se puede programar de forma que cada sujeto trabaje de forma individual o bien en pequeños grupos, haciendo que el sistema informático favorezca la discusión entre los alumnos (Shannon, 1990). Finalmente, diferentes trabajos constatan que la incorporación de las prestaciones que ofrecen los nuevos sistemas multimedia (por ejemplo, la inclusión de material videográfico, sonido y animación) permitirá crear situaciones simuladas muy próximas a las que ocurren en la realidad, lo cual facilitará que los futuros clínicos desarrollen habilidades específicas tales como el reconocimiento y la valoración de diferentes formas de comunicación no verbal en sus futuros pacientes (Garrud, Chapman, Gordon y Herbert, 1993). La disponibilidad de programas de simulación para el entrenamiento también ofrecerá una gran flexibilidad al aprendizaje, ya que en cualquier momento los usuarios podrán detener el proceso objeto de estudio, reiniciarlo e incluso alterarlo.

Y no podemos finalizar esta lista de prestaciones sin destacar una de las que, a nuestro juicio, constituye otra ventaja interesante del uso de la tecnología informática en el ámbito del entrenamiento de los futuros profesionales: la posibilidad de que los usuarios (es decir, los futuros profesionales) desarrollen, junto a las habilidades clínicas requeridas, habilidades en el manejo de los ordenadores, y también que abandonen cualquier actitud de rechazo hacia los sistemas informáticos. Esta consideración es relevante en la medida en que en el futuro estos mismos estudiantes deberán manejar diferentes programas informáticos para las tareas de toma de decisiones clínicas.

Programas de ayuda al diagnóstico y la toma de decisiones clínicas

Uno de los aspectos más relevantes y controvertidos que se ha estado estudiando en los últimos años concierne al uso de los ordenadores como herramienta de ayuda para el diagnóstico y la toma de decisiones clínicas. Tal como hemos visto, esta controversia se inició con Meehl (1954), quien postuló que las predicciones estadísticas eran más consistentes y precisas que las ofrecidas por los clínicos. Posteriormente, se han venido estudiando qué factores podrían explicar por qué los clínicos no producen juicios fiables, y se han destacado tres variables básicas (Kleinmuntz, 1984). En primer lugar, los profesionales humanos están influidos por sesgos aparecidos en sus procesos cognitivos. En segundo lugar, los diagnósticos basados en la intuición no pueden considerarse en modo alguno como productos objetivos de la evaluación. Y en tercer lugar, parece ser que los clínicos tienen problemas para diferenciar qué información es relevante y qué información es irrelevante. En esta misma línea, se ha postulado que cuando se elimina el elemento de interpretación del proceso de evaluación (en este caso el clínico), los diagnósticos son más precisos (Pardeck y Schulte, 1990). Esto implica que la fiabilidad de la evaluación conducida a través del ordenador será notablemente mayor que la evaluación conducida por clínicos humanos (Murphy y Pardeck, 1986), ya que en este caso las decisiones se basan en el análisis de datos obtenidos y combinados a través de procesos consistentes y totalmente estandarizados.

Esta reflexión ha conducido en la actualidad al desarrollo de sistemas complejos de ayuda en el proceso de toma de decisiones clínicas. Debido a la importancia que en un futuro tendrán estas aplicaciones informáticas, les dedicamos un espacio mayor en los siguientes

apartados, donde presentamos dos de los sistemas más conocidos: los sistemas expertos, especialmente diseñados para aprehender el proceso de diagnóstico de los clínicos humanos (Gingerich, 1990), y las redes neuronales. Como se comprobará, estos programas cuentan con múltiples aplicaciones, ya que permiten identificar y clasificar las alteraciones de los sujetos, seleccionar los instrumentos de evaluación más adecuados para cada caso, escoger la estrategia de intervención más apropiada en cada sujeto y evaluar la eficacia de los tratamientos.

2.3.2. Estudio de la equivalencia entre las pruebas tradicionales y su adaptación informatizada

En los apartados anteriores hemos constatado que uno de los primeros usos (y de los más habituales) de los ordenadores en el ámbito de la psicología ha consistido, precisamente, en realizar nuevas versiones informatizadas de las pruebas habituales de evaluación. Esto ha originado un gran número de estudios en torno a diferentes áreas de investigación, de las cuales destacan dos líneas diferentes pero que guardan una estrecha relación: los trabajos que evalúan la actitud de los pacientes hacia esta nueva forma de evaluación y los trabajos que miden el grado de equivalencia de los nuevos protocolos informatizados y las versiones tradicionales de lápiz y papel.

Nos ha parecido interesante abordar estos aspectos porque, en última instancia, los consideramos relevantes para la evaluación de la relación coste-beneficio que implica el proceso de evaluación psicológica informatizada, en especial de todas aquellas variables intermedias que podrían estar modulando dicha relación. De otro lado, también pensamos que este apartado es interesante en la medida en que se acepta que las nuevas tecnologías informáticas complementan y/o sustituyen los tradicionales métodos de evaluación. En este caso, hemos querido plantear y evaluar dos cuestiones básicas: (1) estudio de la equivalencia entre los resultados obtenidos con el formato tradicional y el informatizado, y (2) estudio de la validez de la información recogida con el ordenador.

En los siguientes apartados recogemos las conclusiones generales básicas derivadas de los estudios que miden el grado de aceptación de las tecnologías informáticas como nuevas herramientas de evaluación psicológica, y también de los estudios que miden la equivalencia psicométrica entre las versiones informatizadas de las pruebas de evaluación y las versiones tradicionales.

Debemos matizar que los estudios revisados en los siguientes apartados han evaluado, fundamentalmente, pruebas de evaluación diseñadas para ser autoadministradas con el ordenador. En dichos casos, nuestras conclusiones se dirigen, básicamente, a este tipo de protocolos.

Finalmente, matizar que el estudio de la equivalencia entre las pruebas que se presentan en formato tradicional de lápiz y papel y su adaptación en formato informatizado conlleva importantes dificultades metodológicas (Honaker, 1988). Entre las más importantes destaca el hecho de que los diferentes estudios aparecidos pueden estar evaluando la misma prueba, pero usando diferentes programas informáticos. Puesto que pequeñas diferencias en el interfaz afectan notablemente la forma en que los sujetos responden (Lowgren y Lauren, 1993), es posible que las inconsistencias que evidencian las investigaciones que evalúan la equivalencia en realidad estén denotando diferencias en las características de los programas que se utilizan.

Asimismo, es importante tener presente que al evaluar la calidad de un instrumento de evaluación informatizado no es suficiente que los índices psicométricos clásicos adopten valores aceptables (Erdman, Klein y Greist, 1985). En este sentido, por ejemplo, es preciso que las estrategias de presentación de los resultados y los modos de interacción sujeto-ordenador sean minuciosamente evaluados. En el siguiente capítulo abordamos formalmente

aspectos de calidad de datos adoptando como punto de referencia el uso de sistemas informáticos de gestión de la información.

2.3.2.1. Estudio de la experiencia de los sujetos frente al uso del ordenador

Se han postulado diferentes razones por las cuales los resultados obtenidos con entrevistas o cuestionarios informatizados podrían variar de los obtenidos con los protocolos tradicionales que los pacientes completan con lápiz y papel. Entre las más habituales destaca la actitud de los pacientes frente al ordenador, factor que a su vez podría depender de aspectos tales como el grado de familiaridad de los pacientes con dichas máquinas, el nivel de ansiedad de los pacientes cuando interactúan con el sistema o los requisitos específicos de la recogida de datos en formato informatizado (Burke y Normand, 1987).

El estudio de la forma en que los sujetos experimentan los nuevos formatos de evaluación ha sido denominado como "estudio de la equivalencia experiencial" por autores como Honaker (1988). En términos generales, se trata de comprobar si los sujetos evaluados muestran la misma reacción emocional, motivacional, perceptual y actitudinal frente a ambos formatos. La razón principal para abordar este estudio se debe a que cuando un sujeto reacciona de forma diferente ante un tipo de prueba, esto podría tener un importante efecto sobre la equivalencia psicométrica, y también sobre otros componentes del proceso de evaluación.

Veamos, por lo tanto, cuáles han sido las conclusiones generales referidas en la bibliografía. Los resultados obtenidos en algunos trabajos han sugerido que los sujetos se muestran indiferentes frente al tipo de evaluación (personal "versus" informatizada) (Erdman, Klein, Greist y Skare, 1992). Otros investigadores, sin embargo, han hipotetizado que el ordenador podría contemplarse como un "intruso" dentro del proceso de evaluación, lo cual a su vez deterioraría el "rapport" necesario para conseguir la colaboración y cooperación necesaria de los evaluados (Colby et al., 1989; Matarazzo, 1986). Esta última argumentación es de crucial importancia de cara a la evaluación de los trastornos psicológicos. Pensemos que, en gran medida, la calidad de los datos depende de la cooperación de los sujetos que se estudian; si no se consigue su colaboración no es posible asegurar la fiabilidad de la información que se registra. Sin embargo, aunque en algunos trabajos se haya sugerido la idea de que los sujetos evaluados prefieren la entrevista clásica con el clínico que la entrevista informatizada o la autoadministración de pruebas de evaluación, cada vez existen más evidencias de que la administración "on-line" de los instrumentos de evaluación psicológica es bien aceptada por los sujetos, e incluso en muchos casos resulta el formato de evaluación preferido. Así lo demuestra el que, según nuestro conocimiento, constituye el primer trabajo realizado en esta línea (Lucas, Mullin, Luna y McInroy, 1977), que constata que un 82% de pacientes muestran una actitud claramente favorable al uso de entrevistas informatizadas. Posteriormente, estos resultados se han replicado en otros muchos trabajos (Alhberg, Tuck y Allgulander, 1996; Burda, Starkey y Domínguez, 1991; Dignon, 1997; Farrell, Camplair y McCullough, 1987; Harrell, Honaker, Hetu y Obwager, 1987; Honaker, Harrell y Buffaloe, 1988; Krieger y Dlugosch, 1991; Murrelle, Ainsworth, Bulger y Holliman, 1992; Petrie y Abell, 1994; Plutchik y Karasu, 1991; Rosenman, Levings y Korten, 1997; Rozensky, Honor, Rasinski, Tovian y Herz, 1986; Sawyer et al., 1990; Sawyer, Sarris y Baghurst, 1991). Los sujetos entrevistados no consideran que el ordenador ejerza un efecto negativo en la interacción con el entrevistador, de hecho algunos consideran que la disponibilidad de sistemas informáticos hace que el clínico que los usa parezca menos tenso y "más profesional" (Baker, 1992; Kobak, Reynolds y Greist, 1994). Estos resultados son generalizables tanto a la población adulta como a la población infantil y juvenil. Los trabajos de investigación sugieren que los jóvenes se sienten cómodos con el uso del ordenador, y también que un porcentaje

considerablemente alto de adolescentes prefiere introducir directamente la información diagnóstica al ordenador que ser entrevistado por un clínico, especialmente en aquellos contenidos que esta población considera difícil discutir con un adulto (Kight-Law, Mathisen, Calandra, Evans y Salierno, 1989). En estos trabajos también se ha hipotetizado que en aquellos casos en los que los sujetos adolescentes manifiestan disconformidad con los instrumentos de evaluación informatizados, esta actitud en realidad podría estar expresando una disconformidad ante el contenido de la evaluación, y no ante los procedimientos específicos de recogida de datos.

En relación con el punto anterior, también se ha evaluado en qué grado los sujetos dan más o menos información según el tipo de formato. Aunque algunos estudios indican que las pruebas informatizadas ofrecen puntuaciones inferiores que las pruebas que se completan en lápiz y papel (Lambert, Andrews, Rylee y Skinner, 1987), la mayoría de trabajos realizados sugieren que muchos pacientes dan más información (o más sincera) cuando responden directamente al ordenador, sobre todo en lo que se refiere a contenidos que de forma general la población suele considerar como personales (por ejemplo la vida sexual) o aquellos otros que en la práctica clínica resultan más atípicos y que popularmente se consideran "raros" o poco deseables (Alhberg et al., 1996; Duffy y Waterton, 1984; Erdman, Greist, Gustafson, Taves y Klein, 1987; Ferriter, 1993; Hinkle, Sampson y Radonsky, 1991; Kobak, Greist, Jefferson y Katzelnick, 1996; Kobak, Reynolds y Greist, 1994; Johnston y Walton, 1995; Levine, Ancill y Roberts, 1989; Meszaros, Engelsmann, Meterissian y Kusalic, 1995; Petrie y Abell, 1994; Robinson y West, 1992; Rolnick, Gross, Garrard y Gibson, 1989; Sawyer et al., 1991). Una de las razones por las cuales se podría producir este fenómeno se debería al hecho de que el ordenador nunca responde de forma "emocional" ante las respuestas que los sujetos ofrecen (Colby et al., 1989; Erdman et al., 1985), es decir, nunca se enfada, se cansa, no expresa descontento o sorpresa, y tampoco se aburre. Bien al contrario, los ordenadores pueden programarse para que actúen ante los pacientes de forma educada y amable, y también para que parezcan comprensivos.

Antes de finalizar este apartado quisiéramos destacar, en relación a lo expuesto anteriormente, que los estudios revisados no permiten concluir que el ordenador es siempre el medio preferido por los sujetos a la hora de informar sobre sus problemas. En nuestra opinión, aconsejamos ser extremadamente cautos en este punto y, tal como sugieren algunos autores (French y Beaumont, 1987; Hofer y Green, 1985; Kight-Law et al., 1989; Spinhoven, Labre y Rombouts, 1993), considerar en cada caso todos los elementos relacionados con la administración de las pruebas de evaluación psicológica en formato informatizado: los factores biográficos (sobre todo edad y nivel educativo), el tipo y el grado de alteración psicológica, la forma de registro de las respuestas (teclado, pantallas táctiles, etc.), la forma de presentación de las preguntas, el grado de experiencia previa y la actitud de cada sujeto frente al uso de ordenadores, o incluso la "habilidad informática" de los sujetos. En estos trabajos se indica que es, precisamente, la interacción de estos elementos la que determina la reacción de los sujetos frente al proceso de evaluación, y por lo tanto es importante clarificar cómo y en qué grado lo hacen.

Aunque los resultados obtenidos hasta el momento no son del todo determinantes, en conjunto, estos estudios sugieren que el proceso de evaluación informatizada no siempre resulta adecuado para todos los pacientes psiquiátricos. Así, por ejemplo, Skinner y Pakula (1986) ya apuntaban que los sujetos con nivel cultural más bajo y sin conocimientos previos en informática serían los que presentarían una actitud más desfavorable y unos niveles de ansiedad más altos durante la evaluación. Más adelante, otros trabajos han confirmado que los sujetos más jóvenes, con mayor nivel educativo, con mayor experiencia previa en el uso de ordenadores y con actitud más favorable al inicio en relación a este tipo de evaluación, son los que posteriormente muestran también una actitud más positiva y un mayor disfrute con estos protocolos (Burke, Norman y Raju, 1987; Spinhoven et al., 1993). En cambio, ni el sexo ni el

tipo o el grado de psicopatología parece que guardan relación con la preferencia o la aceptación de los protocolos informatizados². Finalmente, también se ha comprobado que el grado de comprensibilidad y la claridad de presentación de las preguntas son factores más importantes que la forma en que se pide a los sujetos que respondan (por ejemplo, uso del teclado frente a pantallas táctiles) (French y Beaumont, 1987).

Es más, creemos que todavía quedan algunas cuestiones por resolver en el estudio de la equivalencia experiencial. Por ejemplo, ¿qué efectos tiene la falta de equivalencia experiencial en el proceso de evaluación? Un estudio preliminar sugiere que no hay efectos sobre las expectativas de los sujetos respecto al tratamiento (Barron, Daniels y O'Toole, 1987), aunque las dificultades metodológicas de este trabajo limitan su capacidad de generalización. Asimismo, nos podríamos cuestionar cómo y en qué grado afecta la falta de equivalencia experiencial en la equivalencia psicométrica. O también cómo afecta la administración informatizada de un test la reacción que pueda tener un sujeto en el futuro frente a otras formas de administración no informatizadas. En relación a este último punto, podría suceder, por ejemplo, que cuando una prueba de evaluación o una forma particular de registrar los datos genera más ansiedad o rechazo en el sujeto, éste reaccionara posteriormente de forma negativa y manifestara disconformidad o desánimo en las posteriores fases de la evaluación. Y al contrario, cuando un sujeto valora positivamente la evaluación informatizada posteriormente podría manifestar descontento ante otras pruebas administradas en un formato diferente.

Finalmente, otro aspecto relacionado con la equivalencia experiencial ha consistido en evaluar la percepción que tienen los sujetos sobre el tiempo que requiere completar las pruebas. En esta línea, los estudios realizados han mostrado de forma consistente que la administración informatizada se percibe como una forma más rápida para ofrecer la información deseada (Gottschalk, Stein y Shapiro, 1997; Honaker, 1988; Kobak et al., 1996), lo cual siempre suele ser valorado positivamente por los sujetos.

2.3.2.2. Estudio de la equivalencia psicométrica

Se considera que dos pruebas de evaluación son psicométricamente equivalentes cuando demuestran que son paralelas, es decir, cuando producen los mismos resultados, las mismas distribuciones, y correlacionan de la misma forma con medidas obtenidas en otras pruebas o variables. Cuando dos formatos o pruebas de evaluación demuestran ser psicométricamente equivalentes, los valores normativos pueden generalizarse directamente de una a otra. En caso contrario, no es posible realizar esta generalización.

El estudio de la equivalencia psicométrica entre las pruebas informatizadas y las tradicionales ha despertado un considerable interés. Muchos investigadores se plantean como objetivo contrastar la existencia de tales equivalencias (Lewis, 1994; Lukin, Dowd, Plake y Kraft, 1985), aunque parecen no tener en cuenta que pasar de un formato a otro implica diferencias sustanciales en la forma de administración de los reactivos y registro de las respuestas.

En efecto, cuando se informatiza un instrumento de evaluación se modifican tanto los reactivos (al menos en lo que concierne a la forma de presentación) como los requisitos de respuesta. Por ejemplo, es habitual que cuando se usa un ordenador se presenten pocas preguntas a la vez en el monitor (lo normal en los sistemas diseñados para la autoadministración es presentar una única pregunta), de manera que el sujeto evaluado nunca puede saber "a priori" cuál será la duración del mismo. Aunque en un inicio no lo pueda

² Es importante tener presente que en la mayoría de estos estudios se han utilizado muestras de pacientes con trastornos de tipo neurótico. Esto implica que los resultados obtenidos no son generalizables a otras formas o grados de psicopatología (pacientes orgánicos, psicóticos, etc.).

parecer, este factor puede llegar a ser importante; pensemos que cuando el sujeto tiene acceso a todas las preguntas desde el principio, puede llegar a considerar que la prueba es demasiado larga, y por lo tanto, estar tentado de abandonarla.

Estas modificaciones, aunque en muchos casos se pretende que sean irrelevantes a los objetivos del test o protocolo, podrían tener un efecto importante en la forma en que los entrevistados responden (Farrell et al., 1987; McLennan, Jackson y Bellantino, 1988; Skinner y Pakula, 1986). Algunos autores, incluso, sostienen que el formato informatizado podría no reflejar el mismo constructo psicológico que el original (Honaker, 1988; Moreland, 1990b), de manera que resulta del todo inadecuado considerarlos "a priori" como equivalentes y generalizar de forma automática las propiedades psicométricas y los datos normativos. Para ilustrar esta afirmación nos bastará presentar un sencillo ejemplo. Imaginemos que estamos evaluando la capacidad de memoria inmediata de los sujetos. En este caso, las puntuaciones de un test informatizado podrían depender de la habilidad de cada evaluado para entrar las respuestas a través de un teclado, factor más irrelevante en la versión tradicional de lápiz y papel. Este hecho haría que fuera del todo inadecuado generalizar las puntuaciones normativas del formato tradicional, así como tampoco las características psicométricas.

Hemos de señalar que muchos estudios concluyen la existencia de una equivalencia psicométrica entre las pruebas tradicionales y las nuevas versiones en formato informatizado (McLennan et al., 1988; Meszaros et al., 1995; Sawyer et al., 1991). Sin embargo, otro conjunto importante de estudios no concluyen que las pruebas informatizadas sean totalmente equivalentes a nivel psicométrico a las pruebas tradicionales. Es más, en algunos casos, los datos indican que no hay equivalencia. Para corroborar esta afirmación nos basta remitirnos a los resultados obtenidos en el caso de las entrevistas informatizadas. Anteriormente, hemos revisado algunos trabajos de investigación que sugieren que con el formato informatizado se recoge más información que con el formato tradicional, y también que los pacientes que interactúan directamente con el ordenador ofrecen más información de contenidos que popularmente se consideran como personales o de aquellos cuya incidencia clínica es relativamente baja. Evidentemente, estos datos implican que ambos formatos no son equivalentes, y que por lo tanto los resultados obtenidos en ambos tipos de evaluación no deberían interpretarse del mismo modo.

Pero no queremos acabar aquí con nuestra reflexión anterior. Aunque no se sabe en qué forma o medida pueda afectar al proceso de evaluación clínica, lo que sí parece claro es que la información adicional obtenida gracias al formato informatizado es válida y precisa. A pesar de que algunos investigadores sugieren que los síntomas añadidos podrían actuar distrayendo la atención del clínico (ya que cuando éste dispone de síntomas poco prevalentes suele percibirlos como más "anormales" y por lo tanto clínicamente relevantes), por el contrario, parece más lógico pensar que esta información adicional facilita tanto el diagnóstico como la orientación terapéutica. E incluso, podría suceder que con la utilización sistemática de los ordenadores como herramientas de administración de las pruebas de evaluación se comprobara que diversos síntomas aparecen con más frecuencia de lo que se estimaba en algunos trastornos psicológicos.

Pero en a este apartado todavía son muchas las preguntas que quedan por responder. Por ejemplo, muchos trabajos orientados al estudio de la equivalencia psicométrica no evalúan todos los elementos asociados a este concepto general, tales como la validez de criterio. Asimismo, muchas de las investigaciones realizadas se han llevado a cabo en población no clínica, y por lo tanto los resultados no son generalizables a dicha población. Finalmente, todavía queda por determinar cómo afectan la equivalencia variables tales como las diferencias individuales de los informantes y el tipo de interfaz hombre-ordenador.

Estudio de la validez

Es importante matizar que el estudio de la validez de los resultados de la evaluación psicológica con ordenador en muchas ocasiones ha recibido más atención que el de la equivalencia (tanto psicométrica como experiencial). Pero como en los casos anteriores, todavía son necesarios más estudios que midan la validez de muchas de las pruebas informatizadas de evaluación que actualmente se utilizan. Asimismo, sería conveniente dotar a los futuros investigadores en estas áreas de guías a la hora de decidir qué aspectos se deben medir, qué criterios se deben seguir y sobre qué componentes focalizar la atención.

Creemos que es precisamente en este punto donde nos hemos encontrado con otro de los temas apasionantes en relación a los instrumentos informatizados en psicología. En efecto, si no hay acuerdo sobre qué criterios deben guiar el estudio de la validez es porque existen diferentes concepciones respecto al uso de los ordenadores en el proceso de evaluación psicológica. En los siguientes párrafos exponemos dos de las corrientes fundamentales (Honacker y Fowler, 1990).

De un lado, algunos autores sostienen que el ordenador debería poder considerarse como un "sustituto" del clínico (Kleinmuntz, 1991). En este sentido, sería suficiente con llegar a demostrar que los ordenadores ofrecen la misma información y toman las mismas decisiones que los clínicos. De hecho, bajo esta óptica se enfatizan prestaciones como la rapidez y fiabilidad de los nuevos procedimientos informatizados. El ordenador sería algo así como un "gemelo" del clínico, sólo que más rápido y también más consistente (Gottschalk et al., 1997; Kobak et al., 1996). Muchos estudios realizados en esta línea se basan en comparar los resultados obtenidos con el ordenador y los resultados obtenidos por el clínico cuando se estudia una misma muestra de sujetos (Vale y Keller, 1987). Puesto que las metas del nuevo sistema consisten en incrementar la rapidez y fiabilidad de la evaluación, la validación está completa y es aceptable cuando se obtienen medidas de concordancia aceptables (Farrell et al., 1987; Wilkinson y Markus, 1989). Lamentablemente, esta estrategia asume que las interpretaciones del clínico son válidas y correctas, lo cual ha resultado ser muy cuestionable en el ámbito de la salud mental. Por lo tanto, previamente es fundamental demostrar la validez de los juicios del clínico experto.

De otro lado, otro grupo de autores sostiene que el ordenador debe "mejorar" (y no sólo duplicar) el proceso de toma de decisiones del clínico. Bajo esta perspectiva, se enfatiza la capacidad del ordenador para manejar la información, realizar análisis complejos y realizar predicciones actuariales fiables. En este caso es importante llegar a demostrar que el ordenador no sólo es un instrumento más rápido y consistente que el clínico, sino que también es más preciso que el mejor de los expertos. Por lo tanto, muchos estudios orientados en esta línea toman como medida básica la validez de criterio de cada sistema.

Finalmente, es curioso comprobar que aunque todavía no se dispone de los suficientes datos relativos a la validez de los nuevos sistemas informáticos, muchos profesionales y usuarios muestran una alta confianza en los ordenadores, de tal modo que en muchas ocasiones aceptan totalmente los resultados de la evaluación si saben que han estado mediados por el uso de estas herramientas (Honaker, Hector y Harrell, 1986). En el capítulo 3 abordaremos formalmente esta cuestión al hablar de calidad de la gestión de datos.

2.3.3. Tests adaptativos informatizados en evaluación psicológica

En los apartados anteriores hemos revisado cómo fueron adoptados los sistemas informáticos dentro del ámbito de la evaluación psicológica. En su mayoría se trataba de programas "sencillos" o "convencionales" especialmente diseñados para la resolución de tareas específicas de recogida de datos, corrección e interpretación de las pruebas psicológicas más utilizadas que ya existían en papel.

En lo que resta de capítulo presentamos programas más "sofisticados" que también están alcanzando un cierto impacto en el seno de la evaluación psicológica y del psicodiagnóstico. En primer lugar hablaremos de los denominados tests adaptativos informatizados. La exposición de este apartado será bastante concisa, ya que su aplicación al ámbito del psicodiagnóstico todavía no es demasiado notoria. A continuación nos ocuparemos más ampliamente de los sistemas expertos y de las redes neuronales con fines diagnósticos. Estos apartados gozarán de una mayor atención debido a que el uso de estos sistemas se está mostrando muy prometedor en nuestra disciplina.

Los tests adaptativos informatizados (TAI) constituyen una de las innovaciones psicométricas más destacables de los últimos años (Muñiz, 1990). Un TAI no es ni más ni menos que un conjunto de ítems (banco) organizados por un programa de ordenador, el cual selecciona y presenta al sujeto los reactivos que mejor le evalúan. En este sentido, el ordenador escoge un nuevo ítem en función de los aciertos y/o errores que comete el sujeto, y finaliza la sesión de evaluación en el momento en que es capaz de situar la puntuación del evaluado con un mínimo error de medida (Olea y Ponsoda, 1996; Renom, 1996).

Los primeros TAI fueron utilizados por organismos públicos (relacionados con Administración Pública, Educación y Defensa) de Estados Unidos en la década de los 70. Debido a sus características y a su base teórica, estos primeros tests fueron principalmente aplicados en medición de capacidades cognitivas, rendimiento y aptitudes. Posteriormente, y debido a las notables ventajas que ofrece (Jones y Zhing, 1994; Meijer y Sijtma, 1995; Renom, 1993a, 1993b; Wise, 1994; Zen Hsiu, 1994), esta metodología se ha extendido ampliamente por otros países, sobre todo en las áreas de evaluación educativa y certificaciones profesionales. En la actualidad están apareciendo las primeras normativas (A.C.E., 1995). Su impacto, sin embargo, ha sido mucho menor en el campo de la clínica y del psicodiagnóstico, a pesar de las ventajas que supondría la adopción de algunas de las ideas que subyacen a estos instrumentos de evaluación (por ejemplo, la disponibilidad de bancos de ítems con preguntas de diagnóstico para diferentes poblaciones de sujetos).

Una de las estrategias que actualmente está recibiendo especial atención es la "Auto Adaptativa", caracterizada porque cada evaluado escoge el grado y el sentido (fácil-difícil) de la progresión. De este modo, la motivación de los sujetos se incrementa y las percepciones de estrés y frustración disminuyen, ya que disponen de mayor control de la situación de evaluación.

2.3.4. El ordenador como herramienta de ayuda para simular el proceso de diagnóstico: los sistemas expertos

En los apartados anteriores hemos indicado que los primeros programas informáticos aplicados al área de la toma de decisiones y del psicodiagnóstico fueron diseñados para realizar análisis estadísticos y psicométricos de un conjunto de datos. Estos sistemas, a lo sumo, acababan por indicar qué diagnósticos eran más probables aplicando técnicas como los árboles de decisión, el teorema de Bayes o el análisis discriminante.

Posteriormente se desarrollaron programas de ayuda para determinar qué información debería recogerse y con qué técnicas de medida. En este caso, la solución ofrecida por los diferentes sistemas solía depender de los criterios de coste de la prueba, de los posibles daños ocasionados al paciente y del valor discriminante de la información recogida.

Pero en la década de los 70 se inicia una nueva revolución teórica y tecnológica, facilitada por la aplicación de conocimientos derivados de la inteligencia artificial al ámbito del diagnóstico clínico (Lasala, 1994): Bajo esta nueva perspectiva se comienzan a desarrollar los denominados sistemas expertos y los sistemas neuronales, ambos con objeto de facilitar el diagnóstico de los trastornos. En este apartado exponemos la aplicación de sistemas expertos al ámbito del psicodiagnóstico.

Queremos comenzar señalando que el término "experto" posee diferentes connotaciones semánticas cuando se revisa la literatura especializada. Los trabajos sobre la actuación de los "expertos" se organizan en torno a dos líneas bien diferentes. De un lado, existe un volumen considerable de publicaciones referidas al juicio clínico y la toma de decisiones (revisados en los apartados anteriores de este capítulo), que destacan la pobre actuación de los profesionales en tareas clínicas debido (en parte) a los efectos de los sesgos y heurísticos cognitivos asociados a dichas tareas. De otro lado, otras disciplinas también interesadas por aspectos relacionados con la cognición (por ejemplo la psicología cognitiva) contemplan a los "expertos" como sujetos competentes que poseen y utilizan estrategias cognitivas diferentes de las que suelen utilizar los clínicos considerados no expertos (tales como un mayor grado de automatización).

Por supuesto, se han realizado propuestas que han intentado reunir ambos puntos de mira (Edwards, 1992; Shanteau, 1992a). Por ejemplo, Shanteau (1992a) desarrolla lo que denomina como "Teoría de la Competencia del Experto". Esta teoría sugiere que los dos análisis expuestos en el párrafo anterior son correctos, aunque incompletos. Desde el punto de vista del autor, el grado de competencia depende de la disponibilidad de cinco componentes básicos: una base de conocimientos lo suficientemente amplia, las estrategias psicológicas propias de los clínicos considerados expertos, las habilidades cognitivas necesarias para la toma de decisiones, la habilidad para usar de forma apropiada estrategias de decisión y las características de la tarea.

El desarrollo de estas posturas reconciliadoras ha permitido el diseño de nuevas e interesantes tecnologías, los denominados sistemas expertos. Gran parte de estos programas, que se suelen considerar "basados en conocimientos" y que se basan en modelos normativos a la hora de abordar el problema de la toma de decisiones clínicas, contemplan el proceso de diagnóstico de forma similar a la representada en la figura 2.1 (Westmeyer y Hageböck, 1992). Esta imagen sugiere que el ordenador constituye una herramienta de ayuda fundamentalmente en las fases 4, 5 y 6. En efecto, en la actualidad cada vez se tiende más a la recogida de información de forma interactiva con el ordenador (fase 4), por ejemplo presentando las preguntas en un monitor y grabando de forma directa las respuestas a través de un teclado. A continuación, los programas existentes aplican diferentes algoritmos de combinación de datos (fase 5) con el fin de obtener resultados diagnósticos fiables y válidos (fase 6).

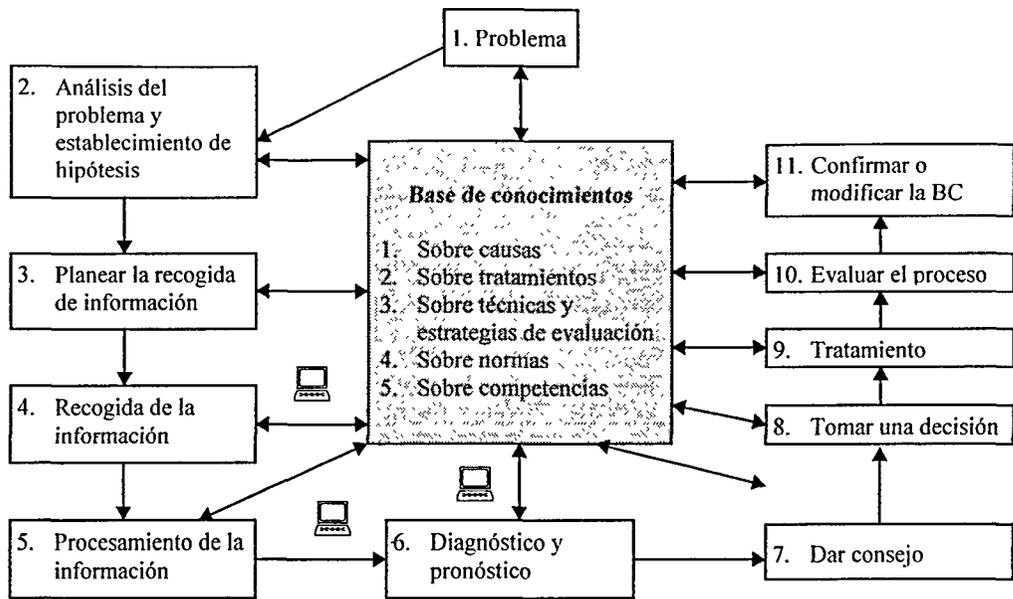


Figura 2.1: Modelo del proceso de evaluación psicológica bajo la óptica normativa (adaptada de Westmeyer y Hageböck, 1992).

Es importante señalar que los problemas planteados en la fase 1 no suelen obtener una solución a partir de la información recabada aplicando un sólo instrumento de evaluación. Por ello, los resultados obtenidos al aplicar una prueba de evaluación nunca deben ser confundidos con el diagnóstico (fase 6).

Asimismo, los sistemas basados en modelos normativos incorporan una base de conocimientos cuyo requisito fundamental consiste en la confirmación empírica. Esto supone que la interpretación "clínica" y la información que se deriva exclusivamente del juicio humano no es suficiente desde el punto de vista normativo. Desde esta perspectiva, lo que se requiere es un proceso de interpretación y predicción automático, pero también "actuarial".

Y es en relación a este punto donde encontramos una de las dificultades fundamentales para implementar de forma estricta la aproximación normativa en el proceso de psicodiagnóstico: la disponibilidad de una base de conocimientos empíricamente contrastada. De hecho, muchos consideran que en la actualidad nuestro cuerpo de conocimientos aparece fragmentado, y en muchos casos no se dispone ni de las reglas necesarias que permiten sistematizar y deducir en cada caso cuáles son los diagnósticos más adecuados, ni tampoco de las distribuciones probabilísticas indispensables para los análisis estadísticos requeridos. En este sentido, si muchas de las consideraciones que realizan los clínicos en el proceso de psicodiagnóstico no poseen confirmación empírica, los modelos adaptados en formato informatizado no asumirán unas directrices estrictamente normativas. Conseguir modelos normativos "estrictos" constituirá uno de los retos de los sistemas futuros basados en conocimientos, aunque en la actualidad este objetivo representa más un ideal que un punto de partida válido en el proceso de evaluación psicodiagnóstica.

Algunos autores establecen una clara diferenciación entre sistemas computerizados expertos y sistemas inteligentes o, si se prefiere, competentes (Godoy, 1996; Kleinmuntz, 1994). En estos trabajos se designa como sistemas "puramente expertos" aquellos que estratégicamente pretenden simular el proceso que siguen los clínicos humanos "expertos", con el fin de llegar a las mismas conclusiones diagnósticas (Edwards, 1992; Hammond, 1992). Por su parte, los sistemas "puramente inteligentes o competentes" constituyen sistemas informáticos que también realizan diferentes tareas de intervención clínica, pero independientemente de la forma de proceder del clínico. En otras palabras, los sistemas

expertos se podrían considerar como alternativas procedentes del trazado de procesos, en la medida que prestan una especial atención a las estrategias que guían la realización de una tarea determinada. Pero los sistemas inteligentes se orientan a la optimización de los resultados, y por lo tanto utilizan el proceso más adecuado en cada caso (pueden recurrir a modelos lineales o a estadística bayesiana, siempre que con ello se consiga una mejora notable en la potencia del sistema). Los conocimientos necesarios para diseñar un sistema presumiblemente inteligente no proceden exclusivamente de la actuación de los profesionales, sino que los diseñadores recurren a diferentes vías para recabar el máximo de información posible (por ejemplo, la consulta recurrente de la literatura científica especializada) y la forma de combinarla. Tal vez por esta razón este enfoque también se ha venido denominando sistemas de producción basados en reglas³ (Frost, 1989).

En este trabajo seguiremos la tendencia que se refleja en un considerable volumen de trabajos, especialmente en la literatura sobre inteligencia artificial, y utilizaremos la denominación de "sistema experto" para designar todo programa que opere de forma eficiente en la realización de tareas complejas (independientemente de si en la forma de proceder simula o no la actuación de los humanos expertos).

2.3.4.1. Cómo se desarrollaron los primeros sistemas expertos y cómo fueron adoptados en el ámbito del psicodiagnóstico

La simulación de los procesos humanos de razonamiento inició su andadura hace ya algunas décadas. Algunos autores sitúan este inicio con los trabajos de Newell y Simon sobre la Máquina de la Teoría Lógica (Newell y Simon, 1956), que probaba teoremas de lógica proposicional. Más adelante, estos mismos autores se interesaron por la simulación mediante ordenador de otras tareas de solución de problemas (Newell y Simon, 1972). Sin embargo, los sistemas expertos de que disponemos en la actualidad se consideran uno de los desarrollos más recientes de la inteligencia artificial. En los siguientes párrafos exponemos brevemente cómo se gestaron y cómo fueron adoptados en el seno del psicodiagnóstico.

En su inicio, uno de los objetivos prioritarios de la inteligencia artificial consistió en el desarrollo de métodos generales para la solución de problemas (Doukidis y Angelides, 1994). Más adelante, los investigadores que trabajaban en este ámbito observaron que la implementación de estrategias generales era insuficiente para la resolución de problemas complejos, y también que el grado en que los sistemas eran capaces de resolver dichas tareas venía estrechamente determinado por la disponibilidad de un cuerpo de conocimientos sobre los aspectos requeridos de la tarea. La idea era bien simple: no existe un "sistema inteligente" universalmente válido; cuando los programas no disponen de información sobre una tarea no pueden resolverla bien, independientemente de lo eficientes que sean en la ejecución de otros problemas.

Este cambio de orientación supuso a su vez un cambio radical en el punto de mira de la inteligencia artificial. El objetivo fundamental consistía en determinar cómo se podían implementar en los ordenadores las bases de conocimientos (las representaciones de los procedimientos y estrategias utilizadas por los "expertos" en la ejecución de un determinado tipo de problema), y los mecanismos que permitían a dichos conocimientos controlar el comportamiento del sistema (reglas). Esta tentativa culminó con el desarrollo de los denominados sistemas expertos basados en conocimientos y/o basados en reglas, considerados

³ Al hablar de sistemas inteligentes y expertos es habitual distinguir tres formas principales de combinar la información y representar el conocimiento: las reglas de producción, los objetos estructurados y la lógica proposicional. En este proyecto nos referimos a los primeros por ser aquellos que han alcanzado de mayor difusión en el ámbito del psicodiagnóstico.

especialmente adecuados para guiar el proceso general de toma de decisiones (Coletti y Regoli, 1992).

Algunos autores consideran que los sistemas expertos constituyeron una auténtica revolución en el seno de la inteligencia artificial. Las razones fundamentales serían dos. De un lado, estos sistemas fueron los primeros productos comerciales nacidos en el seno de la inteligencia artificial. De otro lado, muchos profesionales podían disponer de herramientas de ayuda para solucionar problemas que hasta el momento no eran abordables con el software tradicional.

Los primeros sistemas expertos adaptados al diagnóstico fueron desarrollados en el seno de la medicina. MYCIN -desarrollado para el diagnóstico de enfermedades infecciosas y la prescripción del tratamiento más adecuado- constituye el primer producto elaborado en este ámbito a finales de la década de los 70. Posteriormente aparecieron otros muchos otros sistemas considerados como expertos (Basavappa, Rao y Harish, 1996; Frost, 1989; Godoy, 1996; Korpinen, 1993): INTERNIST, que se usa en medicina interna y se basa en estructuras de objetos (la versión actual se denomina CADUCEUS); VM, que monitoriza a un paciente utilizando un dispositivo de respiración mecánica después de una operación; AI/COAG, para consultas acerca de desórdenes de homeóstasis humanas; DAD, para el diagnóstico de depresión y demencia; OMS, que se basa en la lógica proposicional y cubre la mayoría de los problemas de medicina general; CASNET, especialmente diseñado para el diagnóstico del glaucoma; EXPERT, construido para facilitar el diseño y la comprobación de diferentes modelos de consulta clínica; ONCOCIN, que permite monitorizar el tratamiento de pacientes externos con cáncer; PIP, un programa especialmente útil para el diagnóstico de problemas renales; PUFF, diseñado para el diagnóstico de problemas pulmonares; SLEEP-EXPERT, para el diagnóstico de las alteraciones del sueño; EPILEPSY-EXPERT-SYSTEM, para el diagnóstico de la epilepsia; etc.

El desarrollo de sistemas expertos en psicología fue posterior y alcanzó un menor auge que en medicina, tanto en las áreas de evaluación y diagnóstico, como en las de tratamiento. Como ya hemos argumentado, una de las razones fundamentales podría hallarse en el hecho de que algunos investigadores que trabajan en inteligencia artificial (probablemente por desconocimiento en los recientes avances en nuestra disciplina) consideran que los métodos de psicodiagnóstico están menos estructurados, y que el cuerpo de conocimientos de la psicología en general todavía resulta demasiado vago como para poder implementarlo de forma exitosa con las técnicas actuales disponibles en inteligencia artificial (Adarraga y Zaccagnini, 1992; Binik et al., 1988). También es interesante señalar que en ambas disciplinas todavía se están construyendo prototipos evolutivos, es decir, programas en proceso de desarrollo sometidos a evaluación e implementación de mejoras. En el seno de la psicología los sistemas expertos se encuentran en una fase de diseño más preliminar que en medicina, donde ya se comienza a hablar de sistemas computerizados expertos aplicables a áreas de variada generalidad.

Bajo nuestro punto de vista, también consideramos que para desarrollar sistemas expertos que sean realmente competentes, todavía es necesario resolver problemas de calidad, estructuración y validez de la base de conocimientos psicológicos. No tenemos que ir muy lejos para hallar un ejemplo claro que sirva para corroborar esta afirmación, basta revisar algunos de los trabajos aparecidos en la reciente literatura que aborda aspectos de fiabilidad y validez de uno de los sistemas diagnósticos más utilizados en psicopatología, el DSM (Carson, 1991; Frances et al., 1990; Morey, 1991; Robins, 1989; Spitzer, 1991).

2.3.4.2. Características de los sistemas expertos en psicodiagnóstico

Desde el punto de vista estructural, los sistemas expertos se componen de cinco elementos básicos (Forns, 1993; Frost, 1989; Godoy, 1996; Hand, 1990; Lasala, 1994; Madsen, Reinke, Fehrs y Yolton, 1991), tal como se ilustra en la figura 2.2:

- Una base de datos o conocimientos: está formada por la información contenida en el sistema experto. Es interesante destacar que los conocimientos de la base vienen dados por el experto en la materia, aunque también pueden acumularse los conocimientos derivados del propio sistema.

En esta estructura se incluyen conocimientos de orden declarativos y operativos (procedimentales). Los primeros describen situaciones establecidas o de carácter hipotético (por ejemplo, la definición de un síntoma, como las alucinaciones). Los segundos, también conocidos como base de reglas, constituyen el conjunto de reglas de producción que organizan y sistematizan los conocimientos declarativos.

La información de la base de conocimientos se organiza en estructuras complejas que reflejan la red de interacciones entre los contenidos, se configura por el conocimiento específico en un dominio de aplicación y se formula en el lenguaje propio del dominio. La fórmula de enunciado más usual explicita tanto los efectos como las condiciones de generación, y establece deducciones a través del encadenamiento asociativo de las reglas. El sistema experto requiere también de un "metaconocimiento" respecto a las reglas de producción. Se trataría de algo así como "meta-reglas" que tienen como finalidad saber qué se conoce, discriminar qué importancia tiene cada información, organizar los conocimientos y producir un cambio de representación.

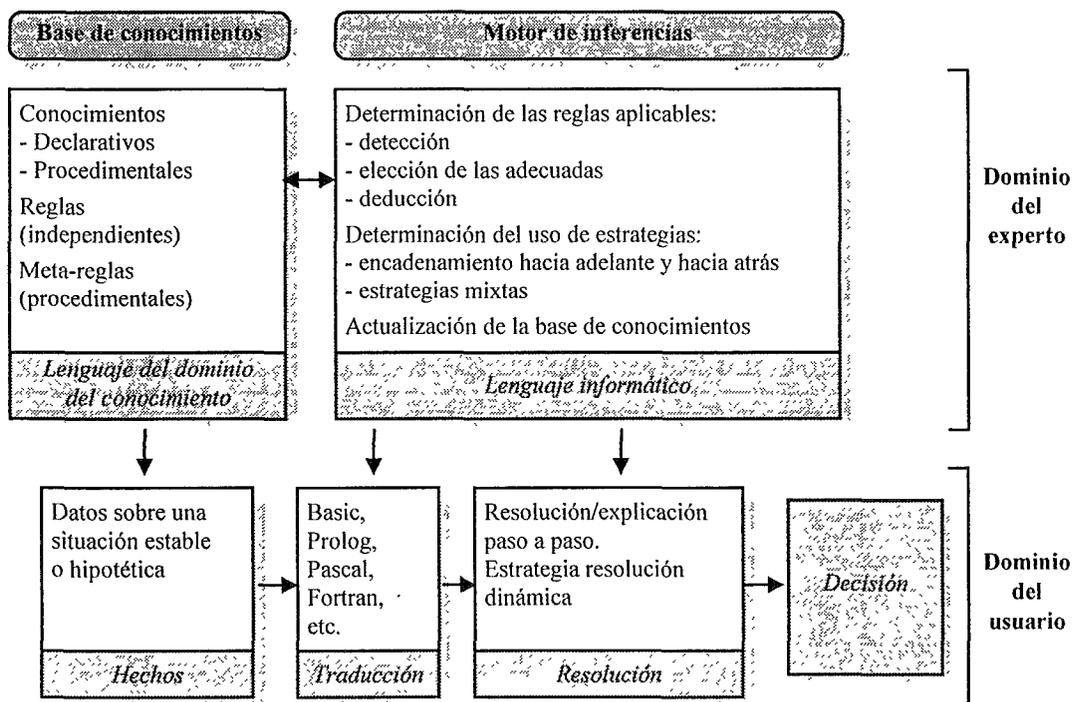


Figura 2.2: Nivel estructural de un sistema experto (adaptada de Forns, 1993).

- Un conjunto de reglas ("de producción"), las cuales representan conocimiento general acerca del dominio del problema. En muchos sistemas expertos basados en conocimientos, cada regla consta de una parte llamada condición *C* y una parte llamada acción *A*, y suelen ser de la forma: Si *C* entonces *A*. Por ejemplo: Si un sujeto ve cosas que otras personas no pueden ver, entonces dicho sujeto tiene alucinaciones visuales. La parte acción de una regla puede significar cosas distintas: la orden de añadir algo a la base de datos acerca del problema que se está considerando, sugerir que el usuario asuma una tarea (por ejemplo evaluar un aspecto determinado del paciente), etc.
- Un motor de inferencias o demostrador de teoremas, encargado de aplicar las reglas que pueden producir cambios y/o adiciones a la base de datos. Se trata de la parte informática del sistema, y puede trabajar con los lenguajes clásicos o con lenguajes específicos para sistemas expertos.

Esta parte del sistema determina el orden y el proceso de aplicación de las reglas destinadas a obtener nuevos hechos a partir de la relación entre los datos del problema y la base de conocimientos establecidos por el experto. El motor de inferencias evalúa y determina si puede poner en marcha el sistema (en cuyo caso explora los conocimientos y elige las correspondientes reglas), dirige y ejecuta las acciones definidas por las reglas (estrategias de encadenamiento de datos y razonamientos) y actualiza la base de conocimientos (acumulación de nuevos datos y generación de nuevas reglas).

Las reglas de inferencia pueden ser de carácter lógico, matemático o mixto. Asimismo, los mecanismos de inferencia utilizados para la formulación de conclusiones y la toma de decisiones pueden ser de carácter deductivo, inductivo o reductivo.

Hay muchas maneras de poder seleccionar módulos de aplicación de reglas y de aplicarlas (Frost, 1989; Lasala, 1994; Madsen, Reinke, Fehrs y Yolton, 1991). El tipo más sencillo actúa de la siguiente manera: (1) evalúa las condiciones de todas las reglas con respecto al estado actual de la base de datos, identificando por tanto un conjunto de reglas aplicables; (2) si no hay reglas aplicables, entonces termina sin éxito; en otro caso, elige arbitrariamente una de las reglas aplicables y ejecuta su parte de acción; y (3) si consigue el objetivo, entonces termina con éxito (en otro caso vuelve al primer punto). Este enfoque, que también se conoce como enfoque sintético, encadenamiento hacia adelante o razonamiento antecedente, se dice que no es determinista debido a la selección arbitraria de una regla aplicable en el segundo paso. Se dice también que el enfoque está "guiado por datos", puesto que el estado de la base de datos es el único identificador de las reglas aplicables.

Un enfoque alternativo a la resolución de problemas en los sistemas de producción es aquel en el que el sistema comienza por el objetivo (expresado por el usuario mediante una expresión cuya certidumbre debe ser determinada) y opera retrocediendo a través de subobjetivos para mostrar cómo se deduce el estado objetivo a partir del estado original. Este enfoque, donde el sistema centra su atención considerando únicamente las reglas que sean relevantes al problema en cuestión, se denomina enfoque "guiado por objetivos" (también denominado enfoque analítico, encadenamiento hacia atrás o razonamiento consciente). En este enfoque, el sistema focaliza su atención únicamente en las reglas que sean relevantes para el problema en curso. Los enfoques guiados por datos y por objetivos pueden combinarse de varias maneras dando lugar a los denominados métodos mixtos.

- La memoria de trabajo. Tiene por objeto almacenar los datos y conocimientos que se manejan en cada momento, y también las conclusiones que se extraen del procesamiento de datos y de los conocimientos.
- Una "interfaz". Es el medio de interacción sistema-usuario, y por consiguiente debe incluir todos los elementos necesarios para permitir y facilitar dicha comunicación. En

este sentido se ha formulado que el usuario debe disponer de los elementos necesarios para entender el proceso de razonamiento del sistema y las demandas que realiza (Rook y Donnell, 1993), independientemente del grado de experiencia previa que posea con los programas informáticos (Montamezi, 1991).

Tal como se ha indicado, a nivel estructural los sistemas expertos se caracterizan por la existencia de una separación entre la base de conocimientos y el motor de inferencias. Desde el punto de vista funcional, aunque cada sistema sigue una forma de actuación particular, algunos de los programas disponibles en psicopatología siguen una estrategia general similar a la resumida en la figura 2.3. En concreto, este esquema representa el funcionamiento del sistema DAI (Adarraga y Zaccagnini, 1992).

Tal como se aprecia en la imagen de la figura 2.3, a nivel funcional algunos sistemas expertos suelen comenzar solicitando información demográfica y psicopatológica de los sujetos. A medida que el usuario entra los datos, los programas pasan la mayor parte del tiempo generando y contrastando diferentes hipótesis diagnósticas (el número de hipótesis activas en cada momento oscila según el sistema). Para comprobar la verosimilitud de una hipótesis concreta se utiliza un criterio de coste, el cual ha sido previamente acordado con el experto. Los sistemas se detienen cuando una hipótesis ha alcanzado un determinado grado de confirmación y no resulta compatible con ninguna otra hipótesis activa, cuando una o más hipótesis han resultado apoyadas y el sistema ha agotado los conocimientos para someterlas a prueba y discriminar entre ellas, cuando todas las hipótesis se han sometido a prueba y no se ha alcanzado el grado de confirmación establecido o cuando entran determinados datos que indican que el sistema no es competente.

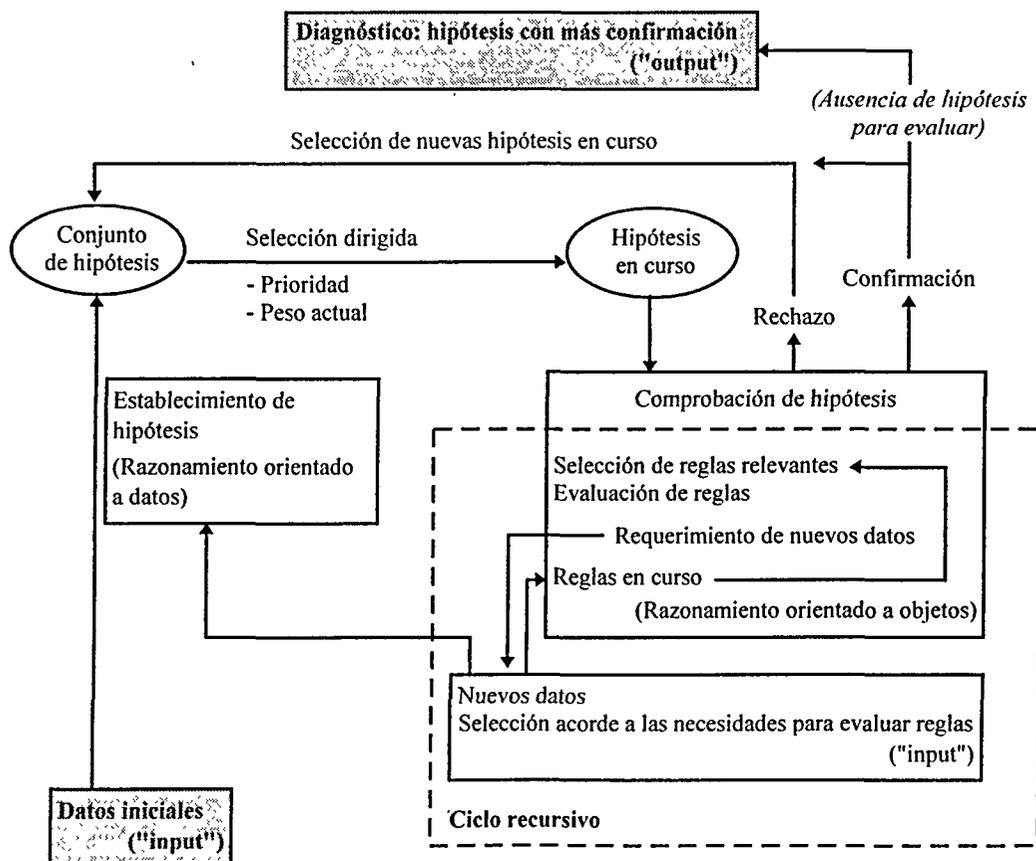


Figura 2.3: Nivel funcional de un sistema experto (Adarraga y Zaccagnini, 1992).

A nivel funcional, los sistemas expertos también cuentan con notables ventajas (Adarraga y Zaccagnini, 1992; Hand, 1990). Una de las características más destacables de estos sistemas consiste en la capacidad para realizar tareas complejas que no son factibles mediante simples algoritmos o modelos estadísticos clásicos, y que de otro modo requerirían la competencia de un humano experto. Por supuesto, las tareas que se abordan deben poder definirse en términos objetivos e intersubjetivos (qué categorías contempla, qué objetivos se pretenden, qué criterios se deben cumplir y qué conocimiento se debe aplicar en cada paso).

Desde su componente funcional, varias de las características más atractivas de los sistemas expertos desarrollados en el ámbito del psicodiagnóstico son (Frost, 1989; Gingerich, 1990; Stefanelli, 1992): (1) la posibilidad de trabajar en condiciones de incertidumbre -cuando falta información los programas son capaces de seguir estrategias diferentes que no requieran de la información omitida, o bien inferirla del resto de datos disponibles-; (2) la capacidad de explicar y justificar sus actuaciones en cada momento (transparencia); (3) la posibilidad de ser adaptados a datos que varían con el tiempo; y (4) la posibilidad de tratar de forma óptima grandes espacios de búsqueda.

Revisemos de forma muy sucinta estos últimos puntos. Por ejemplo, el tema de la incertidumbre ha despertado un gran interés en el ámbito de la investigación en psicopatología. Sabemos que en esta disciplina es habitual que el conocimiento con el que razonan los humanos sea inexacto o incierto de algún modo, debido a factores tales como la ausencia de información, la falta de fiabilidad de la fuente de conocimientos, la ausencia de sistematización en las observaciones o en la formación de juicios, o la ambigüedad en la expresión del propio conocimiento disponible. En este sentido, se han desarrollado varias teorías para dar cabida a la incertidumbre, y muchas de ellas han sido adoptadas por los sistemas expertos basados en reglas de producción. En este sentido, sólo es necesario remitirnos a teorías como (Frost, 1989; Mogre, McLaren, Keller y Krishnapuram, 1994; Perrin, Vaughan, Yadrick y Holden, 1993; Stefanelli, 1992): (1) la teoría de la probabilidad, que proporciona un medio para tratar el conocimiento referido a acontecimientos aleatorios; (2): la teoría de la certidumbre (Shortliffe y Buchanan, 1975), desarrollada para ser utilizada en sistemas expertos en un intento por superar algunos de los problemas tradicionalmente asociados a la teoría de la probabilidad, ya que proporciona un medio de manipulación de estimaciones subjetivas de certidumbre; (3): teoría de la evidencia Dempster/Schafer, que diferencia entre incertidumbre e ignorancia -en lugar de probabilidades se especifican las "funciones de creencias", mediante las cuales se puede poner límites, sobre la asignación de probabilidades, a acontecimientos en lugar de tener que especificar las probabilidades exactamente-, ilustra el efecto de la ignorancia al razonar con un conocimiento incierto y proporciona métodos para calcular funciones de creencias en combinaciones de evidencia (cuando los límites determinan exactamente las probabilidades, esta teoría se reduce a la teoría de las probabilidades, lo cual la ha hecho heredar muchos de sus problemas); (4): teoría de la posibilidad, que sirve para tratar la vaguedad de manera distinta a la aleatoriedad, se basa en trabajos sobre conjuntos borrosos y proporciona reglas para calcular las posibilidades de expresiones que implican variables borrosas o problemas que por su naturaleza se hallan definidos de forma imprecisa; (5): teoría del cálculo de la incidencia, basada en la teoría de la probabilidad en teoría de conjuntos, según la cual la probabilidad de una fórmula se basa en un conjunto de situaciones (interpretaciones o mundos posibles) a los que denomina "incidencias"; y (6): teoría de la plausibilidad, que proporciona líneas orientadoras para razonar con conocimiento obtenido a partir de fuentes que no son totalmente fiables y permite razonar con conjuntos inconsistentes de conocimiento obtenido a partir de fuentes imperfectas.

En cuanto a la adaptación a datos que varían con el tiempo, muchas aplicaciones de sistemas expertos implican situaciones de cambio, por ejemplo en diagnóstico psicológico los síntomas cambian con la progresión de los trastornos y con la terapia que se aplica. Por esta

razón, muchas tareas expertas, especialmente aquellas que implican la planificación y/o la predicción futura, requieren tratar con múltiples futuros posibles. En este sentido, es necesario que el desarrollo de los futuros sistemas expertos contemple esta fuente de variación sintomática.

En relación a los grandes espacios de búsqueda, se sabe que muchas tareas expertas implican el manejo de estos espacios. Entre las distintas estrategias que se han implementado en los sistemas expertos y se han mostrado efectivas en este sentido destacan (Frost, 1989; Lasala, 1994): el razonamiento por eliminación, el empleo de la abstracción, el razonamiento analítico, el principio del compromiso mínimo, el uso de la heurística, las líneas múltiples de razonamiento y el recurso a fuentes múltiples de conocimiento o experiencia.

Finalmente, el aspecto referido de la transparencia guarda estrecha relación con las percepciones del usuario sobre lo razonables que resultan las actuaciones del sistema. Pensemos que es fundamental que todas las explicaciones que ofrece el programa sobre su propia actuación sean completas, claras y razonadas, ya que de otro modo será poco aceptado a pesar de que ofrezca soluciones correctas. En este sentido se ha argumentado que la credibilidad de un sistema experto depende, hasta cierto punto, de su habilidad para explicar su razonamiento.

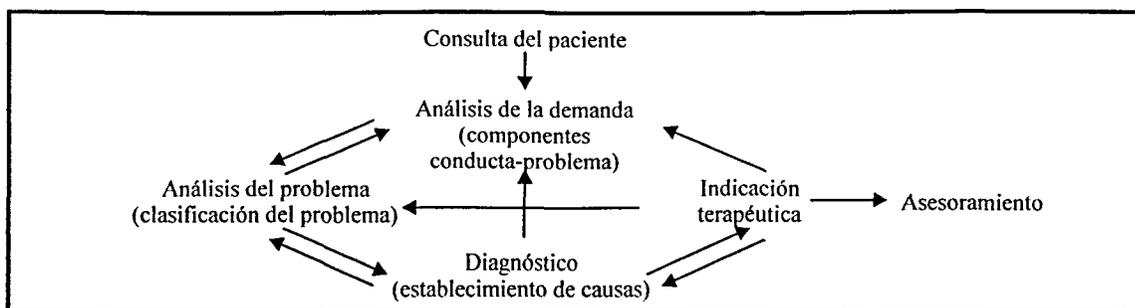
2.3.4.3. Desarrollo de sistemas expertos en psicopatología y psicodiagnóstico

Es importante iniciar este apartado señalando que aunque recientemente han aparecido diversos sistemas de ayuda en el ámbito de la toma de decisiones clínicas y diagnósticas en psicopatología, todavía existen pocos sistemas expertos en este ámbito (sobre todo si tomamos como elemento de comparación la medicina). Por poner algunos ejemplos de los programas más conocidos (dentro y fuera de nuestro país) destacamos:

1. El sistema Psymedia (Hageböck, 1990; Westmeyer y Hageböck, 1992), que constituye un sistema de inferencias que ayuda al evaluador en la tarea de valorar, comparar e interpretar los datos de las pruebas basadas en la teoría clásica de los tests. El objetivo prioritario por el cual fue diseñado consiste en la obtención de información estadística de pruebas psicométricas complejas, independientemente de la forma de administración de dichas pruebas. Uno de los inconvenientes que se ha imputado a este programa es la imposibilidad de manejar e interpretar datos de diferentes pruebas.
2. El sistema DAI (Adarraga y Zaccagnini, 1992), que simula el razonamiento del experto en el diagnóstico del autismo infantil, identificando tanto el patrón de la conducta autista como los diagnósticos apropiados en el caso de decisión diagnóstica de no autismo. La base de conocimientos del programa fue elaborada a partir de la experiencia verbalizada de un psicólogo considerado experto en materia de autismo y otros trastornos psicológicos también severos. Las reglas de la base fueron originalmente escritas en lenguaje natural en colaboración con este profesional, y posteriormente traducidas en lenguaje informático. El sistema fue modificado de forma iterativa según las salidas (resultados) que ofrecía para diferentes grupos de entradas. Este proceso se prolongó hasta que el programa fue capaz de ofrecer las respuestas consideradas "adecuadas" en todos los casos.

En su versión más actual, DAI facilita el diagnóstico de autismo, esquizofrenia infantil, psicosis degenerativa, mutismo selectivo, depresión anaclítica, psicosis simbiótica, síndrome de Rett, síndrome de Asperger, disfasia receptiva severa, retraso mental, trastornos generalizados del desarrollo y trastorno no específico. Asimismo, el programa puede concluir en ausencia de trastorno.

3. El sistema DYSLEXPert (Blonk, Van den Bercken y De Bruyn, 1996), que constituye un sistema experto para el diagnóstico de la dislexia. A diferencia de DAI, en el caso de que la conclusión sea el diagnóstico de dislexia, este sistema ofrece indicaciones terapéuticas.
A nivel funcional, previamente al inicio de una sesión con este sistema, el usuario debe obtener datos sobre el paciente a través de entrevistas o cuestionarios. Una vez proporcionada esta información, el sistema ofrece una primera hipótesis sobre qué áreas del lenguaje, de la psicomotricidad y de la personalidad podrían ser problemáticas. A continuación, el programa también facilita información sobre qué pruebas deberían ser utilizadas para recabar más datos diagnósticos, y qué elementos deberían ser especialmente observados durante esta nueva recogida de información. El programa concluye si se trata o no de dislexia siguiendo reglas heurísticas, y además proporciona información sobre otros diagnósticos posibles tales como disfunción cerebral mínima o trastornos generalizados del desarrollo.
4. El sistema SEXPERT (Binik et al., 1988; Binik, Meana y Sand, 1994; Ochs, Meana, Mah y Binik, 1993; Ochs, Meana, Pare y Mah, 1994), que consiste en un sistema experto para el diagnóstico y la orientación terapéutica de disfunciones sexuales.
El programa ha sido diseñado para efectuar una entrevista interactiva con ambos miembros de la pareja. Éstos únicamente tienen que responder las preguntas que el sistema les formula secuencialmente, y van obteniendo "feedback" sobre los problemas que el programa delimita. Cuando SEXPERT identifica una disfunción, la sesión continúa para intentar ubicarla en un determinado contexto, y también para identificar posibles causas o factores de mantenimiento. Al finalizar la sesión, el programa recomienda un programa de tratamiento, que la pareja puede aceptar, rechazar o modificar.
5. El sistema DIASYS (Diagnostic System, Westmeyer y Hageböck, 1992), que constituye un programa interactivo basado en reglas que se aplica al área de la psicología educativa, y permite realizar el diagnóstico completo de los sujetos. A nivel funcional, y al igual que DAI, el programa DIASYS parte de un conjunto de datos preliminares que previamente han sido recabados sobre el caso actual, y a partir de esta información inicial genera todas las hipótesis posibles que le permite la base de conocimientos. A continuación, el programa indica qué información se debe recoger, y también las pruebas que deberían utilizarse, para continuar el proceso de evaluación y diagnóstico. A partir de esta nueva información administrada al sistema, éste comprueba cada hipótesis.
6. El sistema CONDALS (Van Aarle y van den Bercken, 1992), que facilita el diagnóstico de problemas de lectura y deletreo.
Las bases teóricas y metodológicas CONDALS, al igual que las de DIASYS, se encuentran en el modelo normativo-prescriptivo de diagnóstico propuesto por De Bruyn (1992). Este modelo constituye un esquema basado en dos pilares básicos: el esquema metodológico y empírico que subyace al ciclo del diagnóstico, y el conocimiento procedimental propio del clínico. En este planteamiento, el ciclo diagnóstico constituye la estructura o meta superior que supuestamente dirige la toma de decisiones clínicas. De otro lado, el proceso particular que se sigue conlleva la delimitación de una estructura dinámica compuesta de cuatro elementos básicos, cuya particular combinación representa un escenario distintivo (De Bruyn, 1992):



Asimismo, CONDALS también participa de la metodología del diseño de facetas de Guttman. La aplicación de esta metodología supone la construcción de una definición empírica del fenómeno de interés (problemas de lectura y deletreo) en facetas o atributos observables. El conjunto de todas las combinaciones posibles entre facetas constituye lo que se denomina frase cartográfica, y contiene todas las tareas relevantes y los aspectos comportamentales que se podrían observar durante la evaluación de las disfunciones de lectura y escritura (es decir, los dominios conductuales objeto de la observación).

7. El sistema KRAEPELIN (Roca-Bennasar, García-Mas, Llaneras y Blat, 1991), que facilita el diagnóstico del trastorno obsesivo compulsivo.

Este programa comienza formulando un conjunto de preguntas al sujeto sobre su estado psicopatológico y sobre los posibles factores asociados con su estatus diagnóstico actual. En el caso de que el sistema no disponga de la evidencia suficiente para diagnosticar el trastorno obsesivo compulsivo, sugiere otros posibles diagnósticos relacionados con la patología obsesiva que podrían explicar los síntomas de los pacientes: trastornos fóbicos, alteraciones afectivas o sintomatología psicótica.

8. El sistema EXPERTAL (Ohayon, 1995), que permite predecir futuras conductas de violencia y agresividad en sujetos con evidencia de trastorno mental grave. Los usuarios de este programa destacan su capacidad para formular hipótesis diagnósticas fiables basándose en las respuestas facilitadas por los pacientes, y también de verificar dichas hipótesis en función de la información nueva que se le facilita en cada sesión.

9. El sistema MRE (Mental Retardation Expert; Hile y Desrochers, 1994), que ayuda a los usuarios a que elaboren estrategias terapéuticas para el tratamiento de conductas agresivas y autolesivas en sujetos con diagnóstico de retraso mental o con importantes trastornos generalizados del desarrollo.

Antes de acabar este apartado nos gustaría destacar que muchos clínicos e investigadores ven los árboles de decisión diagnóstica especificados en sistemas como el DSM o la CIE como la estructura fundamental que requiere un sistema experto de decisión diagnóstica. Esto ha originado la aparición de diferentes sistemas, denominados clínicos-expertos, que se basan directamente en estas clasificaciones. Por ejemplo, el DTREE (Raffoul y Lyles, 1993) y el DSM-III-X (Schuler y Tinger, 1991), basados en el DSM-III-R (A.P.A., 1987), permiten establecer un diagnóstico o clasificación en función de los datos que se introducen obtenidos en fases anteriores de la entrevista. Pero debido a los problemas de fiabilidad y validez de estas clasificaciones, diversos trabajos reiteran que previamente al desarrollo de verdaderos sistemas expertos de ayuda taxonómica es preciso que los sujetos "expertos" que elaboran estos sistemas de clasificación lleguen realmente a un acuerdo sobre los contenidos básicos declarativos de cada trastorno, para poder de esta forma redactar las reglas asociativas que determinan las correspondientes decisiones diagnósticas.

2.3.4.4. *Valoración de los sistemas inteligentes y expertos informatizados*

Una de las principales líneas de trabajo en relación al uso del ordenador como herramienta para la simulación del proceso diagnóstico fue la desarrollada por Kleinmuntz (1963a, 1963b, 1984). Los aspectos procedimentales, que constituían una aproximación isomorfa al proceso de toma de decisiones, eran bien simples: se escogían clínicos considerados "expertos" (esto es, con amplia experiencia y adecuada resolución), y se les pedía que completaran unos informes sobre las decisiones que iban tomando a lo largo del proceso de razonamiento clínico. A continuación, estas decisiones se traducían en forma de reglas de producción en los programas de ordenador. Finalmente, se probaba la validez del sistema construido comprobando el grado de concordancia entre los diagnósticos del clínico y los ofrecidos por el sistema.

Tal como se deriva del párrafo anterior, los programas propuestos por Kleinmuntz eran diseñados, al menos de forma aparente, para captar las fases del proceso que siguen los humanos expertos, ya que se basaban en los informes facilitados por dichas personas, respetaban la forma usual del clínico para resolver un problema, operaban paso a paso (de forma secuencial), y representaban la información en un nivel más detallado.

Y es precisamente en este punto donde se encuentran los inconvenientes fundamentales que se han achacado a estos programas informáticos: el grado de verosimilitud acerca de la fidelidad con la que los expertos humanos informan sobre los procesos cognitivos que siguen y las estrategias decisionales que adoptan al enfrentarse a tareas de toma de decisiones (Hoffman, Shadbolt, Burton y Klein, 1995). Asimismo, y puesto que los diseñadores del software suelen añadir elementos no informados por el sujeto para lograr que el sistema que supuestamente simula al clínico funcione correctamente, no se puede considerar que tradicionalmente los profesionales dispusieran de acercamientos que estuvieran basados estrictamente en el trazado del proceso (como la simulación con ordenador de la actuación del clínico) o que reprodujeran fielmente la formación de juicios humanos. En todo caso, los sistemas tenían como objeto ayudar al clínico y mejorar sus resultados corrigiendo las limitaciones y sesgos inherentes a los procesos cognitivos humanos.

En la actualidad el panorama no ha cambiado sustancialmente. Las estrategias de valoración se pueden enfocar desde distintas perspectivas: clasificación, validación, verificación, eficacia, eficiencia y usabilidad (Mitta, 1991; Suen, Grogono, Shinghal y Coallier, 1990). Asimismo, se considera que a la hora de valorar un sistema experto o inteligente se pueden utilizar diferentes criterios, en principio podríamos suponer que tantos como características y aplicaciones se les atribuyan a dichos sistemas (Lee y O'Keefe, 1994; Sharma y Conrath, 1993). En este caso, y con objeto de facilitar el proceso de valoración y verificación, posiblemente la diferenciación entre sistemas expertos e inteligentes sí que permitiría ayudar a establecer cuáles serán los criterios que se deberán contemplar a la hora de comprobar la calidad y las prestaciones de estos nuevos sistemas computerizados. Así, si el sistema ha sido diseñado con objeto de que funcionalmente simule el comportamiento de un humano experto, el criterio de validación constituye la actuación del propio clínico: se trata de comparar al humano y al programa bajo las mismas condiciones (comparación descriptiva). Sin embargo, si el sistema ha sido diseñado con objeto de resolver con éxito una determinada tarea, independientemente de la forma en que habitualmente procede un clínico, el criterio de validación consistirá en el grado de éxito, en la calidad y en la eficiencia con la que se realiza dicha tarea: la comparación es de criterio o normativa (ejecución perfecta).

Entre las mayores dificultades que se ha imputado al uso de los sistemas "propriadamente expertos" destacan (Godoy, 1996; Shanteau, 1992a; Smith y Kida, 1991): (1) puesto que los sistemas informáticos expertos simulan la actuación de estos profesionales, no se puede descartar la posibilidad que comentan estos mismos errores en su forma de proceder; (2) de acuerdo con lo anterior, los sistemas expertos nunca podrán ser soluciones adecuadas frente a

los problemas de razonamiento clínico, y su utilidad quedará limitada a aquellos casos en los que es preciso resolver tareas complejas que requieran altos grados de pericia pero sin la disponibilidad de sujetos humanos expertos; (3) es difícil lograr diseñar un sistema que reproduzca al 100% la forma de actuar del clínico experto (la evaluación de los programas existentes indica que muchas veces lo que se logran reproducir son los resultados y las "apariencias" en cuanto a la forma de actuar, y no tanto la verdadera forma de enfrentar y resolver los problemas); (4) finalmente, algunos autores indican que una de las características de muchos sistemas expertos es la manifiesta "rigidez" que muestran en su funcionamiento, características más propias de los clínicos sin experiencia que de los clínicos expertos.

En cuanto a la valoración de los sistemas expertos en relación a las estrategias actuariales, sabemos que gran parte de los "sistemas inteligentes" en clínica utilizan técnicas estadísticas, fundamentalmente ecuaciones de regresión y formulación bayesiana (Carroll, 1987). Sin embargo, lamentablemente, somos conscientes de que el empleo exclusivo de ciertas fórmulas conlleva un proceso poco flexible en la toma de decisiones, y también puede comportar problemas debido a que no siempre se satisfacen los supuestos teóricos y metodológicos que requieren. En este sentido, los sistemas inteligentes resultan menos rígidos, ya que pueden trabajar y suplir deficiencias o incluso la falta de información (en ocasiones de forma análoga a como actuaría un humano experto). Lo que pretendemos decir es que el uso de los sistemas inteligentes y de las fórmulas no deberían verse como alternativas excluyentes, sino complementarias.

En síntesis, desde el punto de vista de su aplicabilidad, los sistemas expertos constituyen herramientas muy útiles en el proceso de toma de decisiones, especialmente en aquellos contextos en los que existen muchos conocimientos, éstos se organizan de forma compleja, evolucionan con rapidez y no están totalmente estructurados, es decir, en aquellas áreas que no se adaptan bien a los sistemas empleados por la informática clásica (Madsen et al., 1991; Ravetz, 1993). Una de las prestaciones más importantes de estos sistemas consiste en su capacidad de ayuda en aquellas situaciones que requieren de una experiencia humana considerable.

De otro lado, si se acepta lo anterior y se opta por utilizar sistemas inteligentes y/o expertos, el profesional se pregunta qué estrategias estructurales y procedimentales se deberían programar a la hora de diseñarlos. En relación a este punto, diferentes estudios sugieren que ninguna estrategia es superior a otra (al menos a nivel general). Cada una posee sus propias ventajas y sus inconvenientes (Clark, 1990), y es la utilidad potencial o la relación coste-beneficio la que hace que algunas estrategias resulten más aconsejables que otras (Godoy, 1996). Por ejemplo, los acercamientos estadísticos son especialmente adecuados en problemas cuya resolución consista en escoger la mejor alternativa de entre un conjunto bien delimitado, si los datos cumplen las condiciones de aplicación y si se dispone de inferencias probabilísticas. De otro lado, los acercamientos que utilizan reglas de inferencia suelen utilizarse cuando el conocimiento se organiza en torno a reglas o tablas, las clasificaciones son fundamentalmente categoriales y no existe alta dependencia contextual. En cuanto a las reglas de producción, son útiles para representar el conocimiento procedimental en el que si se cumplen un conjunto de condiciones se realiza una determinada acción. Finalmente, los acercamientos de objetos estructurados (muy recientes en el seno del psicodiagnóstico) son especialmente adecuados cuando el conocimiento a representar es fundamentalmente descriptivo, existen dimensiones cuantitativas y cualitativas, existe una gran dependencia contextual y existen muchos resultados potenciales. Este último suele ser el caso de muchas situaciones clínicas, de diagnóstico y de tratamiento, especialmente cuando se utilizan sistemas nosológicos del tipo DSM o el CIE.

Finalmente, en la actualidad parece que los profesionales consideran los sistemas expertos e inteligentes informatizados más como una curiosidad experimental que como herramientas de ayuda en su labor cotidiana (Shanteau, 1992a, 1992b). Sin embargo, los datos

acumulados hasta el momento sugieren que las prestaciones que ofrecen son muy notorias, independientemente del grado de experiencia de los profesionales. Se trata de sistemas muy valiosos en la medida en que permiten acumular, organizar y codificar un gran volumen de conocimiento. Además, a nivel funcional descomponen, formulan y abarcan la resolución de nuevos problemas de forma estandarizada, rápida, fiable y "natural", como si todos ellos tuvieran respuestas igualmente fáciles (Kleinmuntz, 1990, 1991).

2.3.5. Diagnóstico clínico a través de redes neuronales

Hemos visto que desde la óptica de los modelos de representación simbólica (también denominados sistemas basados en el conocimiento) se han realizado aportaciones relevantes al área de la psicopatología, especialmente en el desarrollo de sistemas expertos para la evaluación y el diagnóstico clínico (Adarraga y Zaccagnini, 1992; Westmeyer y Hageböck, 1992).

En este apartado presentamos otros sistemas de evaluación psicológica muy actuales que encuentran su fundamentación teórica en otros modelos, los denominados conexionistas. En un principio, estos modelos hallaron una mayor aplicación en el abordaje de cuestiones teóricas surgidas en el marco de la psicobiología y la psicología básica (Casper, Rothenfluh y Segal, 1992). En la actualidad, uno de los usos más interesantes de estos modelos se encuentra en el desarrollo de redes neuronales, cuyo uso está facilitando resultados también muy fructíferos en el ámbito de la psicología clínica.

Durante las dos últimas décadas, los sistemas basados en redes neuronales han sido aplicados con éxito en distintas áreas de la medicina (Baxt, 1990, 1991; Bassoe, 1995; Bounds, Lloyd y Mathew, 1990; Forsstrom y Dalton, 1995; Madsen y Yolton, 1994; Papa, Stone y Aldrich, 1994; Poli, Cagnoni, Livi, Coppini y Valli, 1991; Riquelme, Zanuto, Murer y Lombardo, 1996). Su uso se ha extendido también al ámbito de la psiquiatría y de la psicología clínica (Aakerlund y Hemmingsen, 1998; Cohen y Servan-Schreiber, 1992; Nair, Reid, y Kashani, 1996; Zou, Shen, Shu y Wang, 1996). En psicodiagnóstico, las redes neuronales se han utilizado en menor grado, aunque cada vez aparecen más trabajos que abordan tareas de clasificación y predicción con estos sistemas. Una de las razones básicas por las cuales se ha afrontado el psicodiagnóstico desde la óptica de las redes neuronales consiste en que los modelos conexionistas muestran una mayor eficacia que muchos sistemas expertos para abordar problemas científicos menos estructurados o que implican valorar situaciones de incertidumbre o con información desconocida (Steyaert, 1994; Forsstrom y Dalton, 1995).

Pero, ¿qué originó realmente el surgimiento de las redes neuronales? Aunque creemos interesante el abordaje de esta cuestión, lo haremos de forma muy concisa. En primer lugar destacaremos que, al contrario de lo que muchos pudieran suponer, la computación neuronal no es un concepto reciente. De hecho, las ideas originales proceden de los años 40, aunque fue necesario aguardar hasta los años 80 para disponer de la infraestructura en "software" y "hardware" requeridos. Durante las últimas décadas, estudiosos interesados en inteligencia artificial (procedentes del área de la psicología cognitiva y de la informática) observaron que ciertos procesos cognitivos podían abordarse en términos abstractos sin hacer referencia a los mecanismos cerebrales que los sustentaban (Ueberla y Jagota, 1993). Este grupo de estudiosos suponían que el cerebro humano constituye, en esencia, un mecanismo computacional altamente sofisticado, y que era posible determinar en qué consiste y cómo funcionan determinados programas informáticos sin necesidad de conocer exactamente la máquina que los controlan.

El enfoque descrito en el párrafo anterior, de carácter básicamente simbólico, permitió la caracterización de aspectos importantes y complejos del proceso cognitivo, tales como la aritmética o la solución de problemas. Sin embargo, esta perspectiva no ha permitido resolver

otros aspectos igualmente importantes referidos a los procesos cognitivos, como el reconocimiento de patrones o el lenguaje. Durante la última década, un grupo importante de estudiosos interesados en área de la cognición humana, comienzan a postular la necesidad de considerar el tipo de "máquina" bajo la cual dichos procesos operan: el cerebro. De este modo se inicia una aproximación de carácter "descriptivo" al comportamiento humano, sustentado en disciplinas como las neurociencias. Esto permitió un notable progreso en el conocimiento del funcionamiento de mecanismos cerebrales básicos, y se establecieron importantes relaciones entre la actuación de determinadas variables neurobiológicas sobre la conducta. Sin embargo, tampoco esta línea de investigación conseguía establecer cómo se traducen los procesos básicos del cerebro en la conducta final que manifiestan los sujetos. En otras palabras, la descripción "pieza a pieza" de la máquina (descripción aislada de los componentes biológicos) tampoco permite explicar el resultado final o las acciones. Como consecuencia, comienzan a emerger nuevas disciplinas de la convergencia entre las neurociencias, la conducta y los procesos cognitivos. Este sería, por ejemplo, el caso de la neuropsicología y también del modelado neuronal. En lo que resta del apartado, abordaremos de forma concisa cómo estos modelos (neuronales) han sido adoptados en el seno de la evaluación psicológica y del psicodiagnóstico.

Tal como su nombre indica, los sistemas neuronales encuentran su origen en un intento por aprehender en formato informatizado algunas de las capacidades cerebrales humanas más sobresalientes (Garson, 1991): el reconocimiento rápido de rasgos-patrones (incluso en presencia de ruido), la capacidad de entendimiento, de interpretación y de manejo de nociones probabilísticas, o la capacidad de asociación y generalización. Hoy sabemos que los humanos resuelven muchas operaciones complejas a través de un aprendizaje basado en ejemplos. Los sistemas neuronales pretenden operar de un modo similar: el sistema se concibe como una "caja negra", a la cual se intenta entrenar facilitándole diferentes patrones que toma como ejemplo de asociación. Dicho de otro modo. Sabemos que el ordenador es más rápido que el cerebro humano en operaciones de cálculo numérico; sin embargo, el cerebro supera ampliamente los sistemas informáticos en otro tipo de tareas que requieren mayor grado de procesamiento y elaboración de la información. Esto ha motivado el intento de modelado del cerebro humano.

Queremos destacar que aunque la motivación inicial que guió el desarrollo de las redes neuronales consistió en el modelado de las funciones cerebrales, muchos de estos sistemas presentan un funcionamiento que no se asemeja totalmente al funcionamiento biológico, ya que no han sido diseñados de modo que detallen los diagramas de circuitos y conexiones característicos de los sistemas biológicos humanos.

2.3.5.1. Los sistemas neuronales a nivel estructural y procedimental

Puesto que no es el objetivo fundamental de este trabajo aplicar redes neuronales con fines diagnósticos, no expondremos ampliamente los componentes estructurales y funcionales. Únicamente indicaremos algunas de las características de los sistemas más utilizados en psicología clínica (una exposición más detallada se puede encontrar en López y Rubio, 1995).

A nivel computacional, las redes neuronales representan un nuevo paradigma, denominado "computación neuronal". En comparación con la computación convencional programada, que requiere que alguien implemente una solución paso a paso para un determinado problema, la red neuronal asume la función de buscar por sí misma y programarse la solución.

Una red neuronal está formada por un conjunto de unidades de procesamiento, denominadas también nodos, que constituyen una representación simbólica de las neuronas biológicas y que están ampliamente interconectados entre sí. Cada neurona está caracterizada

en cualquier instante por un valor numérico denominado valor o estado de activación. Este estado es consecuencia de la suma de señales que recibe la neurona a través de sus conexiones, de los pesos asociados a dichas conexiones y de su umbral. Este mecanismo de ponderación se denomina "función de estado".

En efecto, cada conexión posee una carga denominada peso sináptico (concepto análogo a la fuerza de asociación en la red biológica, y a los coeficientes de los parámetros en los modelos de regresión). Cada nodo posee un determinado número de entradas o "inputs" (representación de las dendritas en la red biológica), que combina mediante un cómputo sumativo hasta que se alcanza un determinado nivel de activación: la entrada total a una unidad constituye la suma ponderada de las entradas que provienen de cada una de las demás unidades de las que recibe conexiones. Las unidades interactúan entre sí transmitiéndose señales. La fuerza de estas señales, y por consiguiente, el grado en que afectan a otras unidades, está determinado por el correspondiente valor o estado de activación.

Las neuronas biológicas sólo producen una salida ("output") cuando el nivel de activación de la célula alcanza un cierto umbral. De forma similar, el nivel de activación interno de una unidad artificial de procesamiento se modifica a través de una función previamente al envío de cualquier signo a otras unidades del sistema. Existen diferentes tipos de funciones de activación o transferencia. Por ejemplo, cuando las salidas de la red son binarias se puede establecer una función "escalón", según la cual la salida de una neurona sólo se activa cuando el estado de activación supera cierto valor umbral. Otra de las reglas comúnmente usadas en los actuales sistemas neuronales es la función de activación logística.

En una red es útil diferenciar entre las diferentes capas. Cada una de ellas encuentra sus entradas en la capa precedente, y sus salidas constituyen las entradas de las capas sucesivas. La capa de entrada está formada por unidades de entrada, caracterizadas por recibir la información de las fuentes externas del sistema. La capa de salida está formada por unidades de salida, que son las encargadas de enviar señales fuera del sistema. Entre las capas de entrada y de salida existen diferentes capas ocultas, cuyas unidades reciben sus entradas y emiten sus salidas dentro del propio sistema, es decir, son invisibles para otros sistemas y sólo sirven para caracterizar las representaciones internas de la red.

El número de unidades que formarán parte de las capas de entrada y salida se suele determinar por el número de datos disponibles en el estudio. En el caso más simple, existiría un nodo en la capa de entrada para cada dato entrante. Para redes cuyo objetivo consiste en la toma de decisiones, debe existir al menos un nodo de salida para cada categoría de respuesta. Para redes de predicción, debe existir un nodo para cada valor que formará parte de las opciones de predicción.

El número de capas ocultas y el número de unidades que constituirán cada una de ellas depende de las características y relaciones entre los datos que manejará la red, así como también el tipo de tarea para la cual el sistema será entrenado. Se ha demostrado matemáticamente que cualquier problema puede ser abordado y representado en redes con al menos dos capas ocultas si se utilizan funciones de activación no lineales (Warner y Misra, 1996).

Finalmente las neuronas de las capas ocultas pueden estar interconectadas de distinto modo, lo que determina las distintas tipologías de redes neuronales. Cuando ninguna salida de las neuronas es entrada de ningún nodo del mismo nivel o de niveles precedentes, la red se describe como de "propagación hacia adelante" o "feedforward". En este caso, la información fluye en una sola dirección, y no hay bucles de retroalimentación. Un ejemplo de red "feedforward" es la Multi Layer Perceptron. Cuando la señal de salida de un nodo puede ser la entrada de otro nodo del mismo nivel o de la misma neurona, se habla de conexiones laterales y autorrecurrentes.

Sabemos que el cerebro humano aprende a partir de las adaptaciones de las conexiones sinápticas. Del mismo modo, los pesos sinápticos de una red neuronal se ajustan para

solucionar el problema que se le presenta a la red. El proceso por el cual se hallan los valores de estas cargas se denomina aprendizaje o entrenamiento. Un criterio utilizado para diferenciar entre las distintas reglas de aprendizaje consiste en considerar si la red puede aprender durante su funcionamiento habitual (redes adaptativas) o si es preciso desconectar la red (inhabilitarla hasta que el proceso finalice). Otro criterio para clasificar las reglas de aprendizaje utilizadas en el entrenamiento de redes neuronales se basa en determinar si existe un agente externo (supervisor); en este caso se distingue entre redes con aprendizaje supervisado y no supervisado.

El entrenamiento de las redes con aprendizaje supervisado consiste en proporcionarle al sistema diferentes ejemplos (de entrenamiento), cada uno de los cuales posee un patrón de datos entrantes asociados a una salida deseada. Cada patrón de entrada produce una salida en la red, la cual se compara con la salida correcta para los datos entrantes. En este momento, la red modificará su representación interna de los datos en función de la diferencia entre los valores deseados y los obtenidos en la salida (corrección de error), procurando en cada caso aprehender las características fundamentales de la información que maneja. Existen diferentes algoritmos para el entrenamiento supervisado de redes neuronales; la retropropagación (procedimiento iterativo también denominado regla generalizada delta) es uno de los más comunes. Otros algoritmos de aprendizaje supervisado son el de refuerzo y el estocástico.

En el aprendizaje no supervisado no se dispone de valores finales para cada patrón de entrada en la red. En este caso, el sistema no dispone de "feedback" durante el proceso de entrenamiento, y crea su propia representación de los datos que maneja. La red aprende agrupando los patrones de entrada según las características comunes que presentan, procedimiento similar al que se utiliza con técnicas estadísticas clásicas como el análisis factorial. Existen dos procedimientos básicos de aprendizaje: el competitivo (las neuronas de cada capa están interconectadas, y cuando una produce una salida alta se inhiben las salidas de las otras neuronas) y el cooperativo (las neuronas de cada grupo trabajan juntas para reforzar sus salidas).

Muchos de las redes neuronales que se emplean en psicopatología se basan en un procedimiento de aprendizaje supervisado. Es habitual que estos sistemas utilicen el algoritmo de retropropagación para modificar los pesos de las conexiones entre unidades, y que midan el aprendizaje de la red a través del cómputo del error RMS (Rumelhart, Hinton y Williams, 1986). La elección de las funciones de activación, del valor del parámetro definido para el entrenamiento y de los ciclos de entrenamiento, puede ser diferente en cada red. Es habitual que esta elección se efectúe en función de los resultados obtenidos en simulaciones previas que dieron lugar a resultados óptimos utilizando dichos valores y funciones (Pitarque, Ruiz y Algarabel, 1995).

2.3.5.2. Uso de sistemas neuronales en psicología clínica

Los resultados obtenidos en psicología clínica muestran claramente que los modelos de redes neuronales pueden ser utilizados como sistemas facilitadores del diagnóstico clínico (Pitarque, Ruiz, Fuentes, Martínez y García-Merita, 1997), tal como ocurre tanto en medicina como en las otras ciencias donde dichos modelos han sido aplicados. Es habitual que la proporción media de aciertos supere ampliamente el 90%, y los tests estadísticos destacan el gran poder predictivo de estos modelos.

Las redes neuronales se han utilizado en una amplia gama de situaciones relacionadas con el psicodiagnóstico, fundamentalmente en problemas que requieren abordar tareas de clasificación y/o predicción. Desde el ámbito de la investigación, estos problemas se habían venido abordando con procedimientos estadísticos clásicos, fundamentalmente mediante modelos de regresión lineal, regresión logística, análisis discriminante o modelos bayesianos.

En cuanto al uso de redes neuronales en psicopatología, conviene resaltar que estos sistemas no sólo facilitan tareas de clasificación y predicción. Trabajos recientes aportan otros usos también muy interesantes de las redes. Planteamos un simple ejemplo: la identificación, entre un conjunto de factores, de cuáles son los que se relacionan con mayor intensidad con las respuestas. O lo que es igual, identificar qué variables modulan, y en qué grado lo hacen, la adaptación psicológica de los sujetos (Kashani, Nair, Rao, Nair y Reid, 1996; Reid, Nair, Mistry y Beithman, 1996). Estos trabajos enseñan cómo medir el cambio que se produciría en las respuestas cuando se van modificando los valores de cada factor concreto (a la vez que se mantienen constantes los valores de los otros factores), una vez entrenada la red. De este modo se establece la significación clínica de las variables evaluadas, o lo que es lo mismo, qué combinación de entradas resulta más significativa a nivel práctico y/o clínico. Como se podrá deducir fácilmente, esto permite la identificación de factores de riesgo, así como también el establecimiento de propuestas terapéuticas y preventivas eficaces dirigidas al control de dichos factores predisponentes a la psicopatología.

Gran parte de los trabajos revisados que utilizan sistemas basados en redes neuronales no pretenden mostrar esta estrategia como una herramienta final y definitiva de diagnóstico (sobre todo porque en su mayoría trabajan sólo con unas cuantas categorías clínicas amplias), sino mostrar que pueden resultar muy útiles para abordar este tipo de problemas (Pitarque et al., 1997). En este sentido, los usuarios de redes neuronales que trabajan en el ámbito del psicodiagnóstico señalan que son necesarios trabajos futuros que aborden el entrenamiento de redes para que aprendan a discriminar entre subcategorías concretas, y no sólo entre las grandes categorías clínicas. El reentrenamiento de la red, favorecido por el continuo incremento de información que se le proporcionaría al sistema como consecuencia de la incorporación de nuevos pacientes, facilitaría diagnósticos más específicos y precisos.

2.3.5.3. Los sistemas neuronales frente a modelos estadísticos tradicionales

¿Qué ventajas e inconvenientes presenta el uso de redes neuronales en relación a los modelos estadísticos tradicionales? Intentemos dar una respuesta breve a esta cuestión. Diversos autores han llevado a cabo estudios de comparación entre los métodos estadísticos y las redes neuronales (Buntine y Weigend, 1991; Garson, 1991; Hruschka, 1993; Ripley, 1994; Warner y Misra, 1996). En sus conclusiones destacan que el uso de los métodos estadísticos comporta el cumplimiento de unas determinadas condiciones de aplicación. Por ejemplo, la aplicación de la regresión múltiple requiere la combinación lineal de las variables independientes y dependiente del estudio, así como también la ausencia de interacción entre las variables predictoras. A medida que se vulneran estos supuestos, se cuestiona la interpretación de los resultados obtenidos con estas técnicas. Las redes neuronales no imponen como condición de aplicación un tipo concreto de relación funcional entre las variables, al contrario, favorecen que ésta emerja de los propios datos. Y es aquí donde radica uno de los valores potenciales de las redes neuronales: se pueden aplicar cuando se desconoce la relación funcional entre las variables independientes y las dependientes, cuando los problemas son complejos, cuando la relación entre los datos no es lineal o cuando existe información desconocida.

Pero si vamos un poco más allá, Warner y Misra (1996) establecen que las redes neuronales constituyen algo así como un abordaje "no paramétrico" del análisis de datos, un análisis robusto en la medida en que permiten resolver problemas complejos cuya distribución subyacente suele ser desconocida. Además, las redes poseen la notable ventaja de la generalización y la adaptación del modelo en respuesta a nuevos conjuntos de datos. Sin embargo, los abordajes "paramétricos" (tales como la regresión lineal) pueden constituir la estrategia de elección cuando el tamaño de las muestras es muy pequeño, y también en

aquellos casos en los cuales se conoce la relación funcional que subyace. Pensemos que las redes neuronales no siempre permiten conocer cuál es la relación subyacente entre los datos, lo cual puede crear desconcierto o incertidumbre al intentar establecer las razones por las cuales la red realiza una determinada predicción.

2.3.5.4. Programas informáticos convencionales, sistemas expertos y redes neuronales

En este apartado, que finaliza la presentación sobre programas de simulación, se establecen unas últimas consideraciones en relación al uso de ordenadores con fines psicodiagnósticos. En concreto, se comparan los sistemas convencionales, los sistemas expertos y las redes neuronales en aspectos meramente computacionales y funcionales.

Nivel computacional

Una de las características más llamativas de los ordenadores convencionales es el almacenamiento de cada fragmento de información en un área concreta de memoria. Esta forma de gestionar la información hace que los programas fallen siempre que algún fragmento de memoria se dañe.

En una red neuronal, la memoria se distribuye a través de los pesos de conexión, los cuales forman una representación interna del problema en el que ha sido entrenado dicho sistema. Esto supone que la misma unidad informativa está contenida y representada en múltiples pesos y puntos de conexión. Este modo particular de gestionar y distribuir la información conlleva una notable ventaja, en ocasiones no muy explotada: aunque algunos elementos de procesamiento de información fallen, o aunque se alteren los pesos de conexión, el comportamiento global del sistema sólo se ve parcialmente alterado (Losilla y Navarro, 1998). Evidentemente, a medida que fallan más elementos de procesamiento, la ejecución conjunta de la red se altera. Pero el sistema ve afectado su rendimiento con bastante lentitud, de tal forma que se protege de los denominados "errores catastróficos". Esta capacidad de "tolerancia a fallos" permite que la red continúe realizando su función (con cierta degradación) aunque se destruya parte de la red (por supuesto, cuando la red neuronal se implemente en un ordenador convencional, fallará de la misma forma que los programas habituales si el procesador central se lastima).

Nivel estructural y funcional

Los ordenadores convencionales poseen en un único programa tanto el conjunto de la información como las instrucciones. Una diferencia fundamental entre los sistemas informáticos clásicos y los sistemas expertos reside precisamente en que éstos últimos incorporan de forma independiente las reglas de producción. Esto permite que los creadores de software inviertan gran cantidad de esfuerzos en el diseño de un programa general que procese la información de formas muy diversas (con el fin de resolver diferentes tipos de tareas), y que la base de conocimientos pueda modificarse y/o incrementarse sin que sea necesario cambiar simultáneamente el motor de inferencias. De este modo, el programa "aprendería" a partir de su propia experiencia, ya que sería capaz de incrementar y reestructurar por sí mismo la base de conocimientos. Y, al menos a nivel teórico, gracias a la redacción asociativa es posible añadir, quitar o variar reglas de un sistema experto sin necesidad de revisar todo el conjunto, lo cual hace de estos programas herramientas en cierto modo flexibles y fáciles de utilizar.

Por otro lado, las redes neuronales generan sus propias reglas en función de la evaluación de los datos que se le proporcionan y de la tarea a la que se enfrenta. Estos sistemas aprenden sobre un problema utilizando reglas de aprendizaje que modifican los pesos de conexión en respuesta a los ejemplos que le proporcionan los datos de entrada y las salidas deseadas. En este sentido, la gran ventaja de los sistemas expertos radica precisamente en el hecho de que los conocimientos y las reglas que se utilizan para llegar a unas conclusiones son siempre totalmente explícitas, cosa que no sucede cuando se emplean redes neuronales.

Los sistemas expertos aplican un conjunto de reglas al análisis de conjuntos de datos. Muchas veces se supone que dichas reglas reflejan el cuerpo de conocimientos y las reglas de producción de los profesionales expertos que realizan tareas que implican la toma analítica de decisiones. Debido a la implementación de estas reglas, es posible establecer qué proceso siguen estos sistemas cuando emiten sus propios juicios. Sin embargo, no es fácil diseñar un sistema experto, ya que muchas veces las reglas de inferencias son desconocidas o la información de la que dispone el clínico experto es difícil de operativizar. Asimismo, cuando en la práctica se programa un sistema experto suele ser difícil adaptarlo para la solución de otros problemas distintos a los que originaron su diseño (en ocasiones, cambiar las reglas de inferencia puede implicar la creación de un sistema experto totalmente nuevo).

Por su parte, las redes neuronales no utilizan reglas de producción rígidas. Estos sistemas aprenden de las entradas que se le ofrecen, y crean sus propias decisiones basándose en el conjunto de patrones que manejan. Esto hace que una red pueda adaptarse para resolver diferentes problemas con bastante más flexibilidad y facilidad.

Algunas redes neuronales poseen, además, otra propiedad de memoria denominada asociación. En una red asociativa la información puede ser elaborada aunque sólo se disponga de un patrón de entrada parcial. Esta propiedad permite al sistema generar respuestas razonables aunque se enfrente a información incompleta o que nunca antes había manejado.

2.4. CONCLUSIONES DEL CAPÍTULO

2.4.1. Los clínicos, aunque "expertos", cometen errores cuando gestionan los datos y cuando realizan valoraciones psicodiagnósticas

La psicopatología en general, y el psicodiagnóstico en particular, constituyen disciplinas que se caracterizan por tratar con problemas clínicos complejos que requieren que los decisores definan operacionalmente, recaben y manejen dos tipos básicos de datos: los que constituyen propiamente el problema y los que se utilizarán después para solucionarlo. Puesto que en la mayoría de los casos existe una ingente gama de posibilidades que podrían estar actuando como problemas clínicos potenciales y como alternativas de resolución, se hace del todo imprescindible evaluar el tipo de actuación que lleva a cabo el clínico (especialmente las variables que se relacionan con el éxito-fracaso en dicha actuación) y los procedimientos a los que puede recurrir para determinar qué datos son los relevantes y en qué modo debe recogerlos y tratarlos.

En relación al primer punto anterior, los resultados obtenidos en numerosos trabajos de investigación han creado un clima pesimista sobre las estrategias procedimentales (fundamentalmente cognitivas) que llevan a cabo los profesionales cuando emiten juicios clínicos. Por ejemplo, entre las conclusiones que se derivan de algunas publicaciones aparecidas en el seno de la teoría conductual de la decisión, del análisis decisional, de la teoría de formación del juicio clínico, de la teoría de solución de problemas y de otras disciplinas también relacionadas con el razonamiento diagnóstico, se evidencia que los profesionales del ámbito de la salud mental cometen errores y falacias cuando se enfrentan a tareas de evaluación y diagnóstico: su actuación no es tan diferente al resto de los humanos cuando

realizan tareas cognitivas cuyo resultado consiste en la emisión de juicios y/o en la toma de decisiones clínicas. Sintetizando algunos de los resultados más destacados obtenidos en numerosos trabajos de investigación se constata que (Adelman y Bresnick, 1992; Arkes, 1991; Arocha, Patel y Patel, 1993; Block y Harper, 1991; Christensen-Szalanski y Willham, 1991; Edwards, 1992; Ellis, Robbins, Schult, Ladany y Banker, 1990; Elstein et al., 1990; Fischhoff y Beyth-Marom, 1988; Garb, 1989; Gigerenker, 1991a, 1991b; Gigerenker y Hug, 1992; Gigerenker, Hoffrage y Kleinbölting, 1991; Godoy, 1996; Goldstein, 1990; Hawkins y Hastie, 1990; Harvey, 1992; Johnson, Grazioli, Jamal y Zualkernan, 1992; Joseph y Patel, 1990; Kahneman, 1991; Keren, 1992; Koehler, 1991, 1993, 1994; Miller y Karoni, 1996; Reilly y Doherty, 1992; Paese y Sniezek, 1991; Smith y Kida, 1991): (1) los profesionales forman hipótesis sobre las que elegir a partir de un conjunto limitado de datos; (2) es habitual que el profesional atienda de forma selectiva a unidades simples de información, frecuentemente las que le resultan más sobresalientes; (3) los profesionales tienden a generar una opinión general de los casos a partir de las primeras hipótesis, y a partir de entonces tienden a guiar su actuación con objeto de confirmar dicha opinión (más que a contrastarla) -por ejemplo procesando mejor aquella información congruente con la hipótesis de trabajo que formulan y subestimando el resto de datos por ser considerados inconsistentes-; (4) también es frecuente que ante problemas con información más inconsistente los profesionales procedan a la recogida de un menor número de datos relevantes, y también que formulen un número menor de hipótesis; (5) aunque los humanos tienden a generar hipótesis ante un conjunto limitado de datos, también es habitual que los profesionales en el ámbito de la salud mental reúnan al final del proceso de evaluación una gran cantidad de información (a pesar del incremento en los costes del diagnóstico y de la sobrecarga que conllevan tales excesos en las capacidades humanas cognitivas de procesamiento de la información); (6) aunque los humanos no están acostumbrados a pensar en términos probabilísticos y cometen errores al evaluar con qué probabilidad previa se presentan los sucesos en la vida real, durante el proceso de generación de hipótesis los profesionales suelen guiarse de la probabilidad "subjéctiva" que poseen en relación a los trastornos; (7) durante la determinación de la probabilidad "final" que corresponde a una hipótesis, es habitual que los sujetos se vean afectados por un fenómeno que en su día Chapman y Chapman (1969) denominaron como "correlación ilusoria" (sesgos producidos como consecuencia de las creencias y/o expectativas previas de los sujetos); (8) los profesionales cometen errores al revisar su opinión frente a la nueva información disponible, ya que muchas veces se centran exclusivamente en los datos que consideran más destacados (por ejemplo porque resultan poco frecuentes); (9) en ocasiones los profesionales adoptan la hipótesis con mayor número de datos a su favor, sin tener en cuenta la información negativa o la magnitud de las diferencias; (10) muchos profesionales, sobre todo los menos experimentados, tienden a buscar hipótesis diagnósticas poco frecuentes, pero llamativas; (11) los profesionales no suelen tener presente la existencia de falsos positivos, y suelen suponer que cuando una conducta se asocia con alta frecuencia a un determinado trastorno dicha conducta resulta predictiva o patognomónica de dicha alteración; (12) bajo ciertas condiciones, los profesionales llegan a diferentes conclusiones clínicas en función de la secuencia en que reciben un mismo conjunto de datos; (13) en lo que se refiere al grado de experiencia, parece que si se toma esta variable de forma aislada no resulta relevante sobre el grado de precisión y la calidad de los resultados del psicodiagnóstico; (14) el grado de accesibilidad y la utilización del sistema de conocimientos, sin embargo, sí parecen depender de variables tales como la necesidad (repetición) con la que el profesional ha necesitado recurrir a una explicación concreta de dicha base de conocimientos; (15) siempre que pueden, los profesionales tienden a formular las conclusiones de un caso en términos de una sola hipótesis (la que mejor explica el conjunto de síntomas); (16) la actuación de un profesional frente a un problema particular no permite predecir cómo será su actuación frente a otro problema diferente, aunque también relacionado con la evaluación diagnóstica (esto es, no

existe una "consistencia individual" de los sujetos en cuanto a la generación de las hipótesis correctas); (17) muchos profesionales continúan realizando tareas de juicio erróneas porque confían excesivamente en la calidad de sus propias respuestas, y las sobrevaloran; y (18) finalmente, la sobreconfianza (identificada por Oskamp, 1965) se ha relacionado con la desestimación de información que podría ser muy relevante en el proceso de diagnóstico, especialmente cuando los datos que no son congruentes con el propio juicio.

En relación con las limitaciones anteriores, parece que muchos errores de diagnóstico no son consecuencia de la recogida escasa de información, sino de la valoración y combinación inadecuada de los datos. De hecho, gran parte de los problemas evidenciados en el estudio del proceso de diagnóstico se relacionan con la incapacidad de los sujetos para procesar gran cantidad de datos, sobre todo cuando éstos son de carácter probabilístico (situación habitual en nuestra disciplina).

Asimismo, una de las causas de tan frecuentes y variados errores en las tareas habituales de evaluación y toma de decisiones clínicas se encuentra en el hecho de que los profesionales que no cuentan con otras fuentes de ayuda externa se guían por reglas o heurísticos siempre que se enfrentan a situaciones complejas que requieren procesar un volumen de datos más o menos considerable (Block y Harper, 1991; Gigerenker, 1991a, 1991b; Gigerenker y Hug, 1992; Gigerenker, Hoffrage y Kleinbölting, 1991; Keren, 1992). La causa por la cual recurren a dichas estrategias conceptuales es bien simple: las capacidades cognitivas humanas son limitadas, en especial las de la memoria de trabajo. Fischhoff y Beyth-Marom (1988) señalan, a su vez, que el uso de heurísticos conlleva la inclusión de errores sistemáticos en las tareas de toma de decisiones.

2.4.2. La introducción de herramientas informáticas mejora la gestión de datos clínicos y la calidad de los diagnósticos

En vista de los resultados expuestos en el apartado anterior, parece evidente la necesidad de adoptar estrategias que ayuden al profesional a efectuar con eficacia y eficiencia las tareas de evaluación psicodiagnóstica. En este sentido, en algunos trabajos se proponen programas de entrenamiento para que el profesional aprenda el uso sistemático de algunos heurísticos que le permitan mejorar los procesos de solución de problemas y adecuar sus diagnósticos a la realidad clínica de los pacientes (Elstein et al., 1990). Este entrenamiento iría dirigido a la optimización de aspectos relacionados con la recogida de la información, la combinación de datos, la generación y la comprobación de hipótesis.

Otros autores proponen, además de dichos entrenamientos, la introducción de sistemas informatizados de ayuda para la evaluación de los trastornos psicológicos y de sus correlatos (Kleinmuntz, 1990). Su argumentación es bien simple. Después de diseñar un programa informático "adecuado" para evaluar los aspectos psicopatológicos, éste presentará un comportamiento estándar que no variará en el curso del tiempo. En este sentido, los sistemas informáticos necesariamente ofrecerán diagnósticos más objetivos que el profesional, ya que funcionalmente sólo operarán en base a la información diagnóstica de la que disponen (en ningún modo se verán influidos por impresiones personales o teorías implícitas sobre la estructura de la personalidad o la configuración de los trastornos), y presentarán una actuación invariante que no se verá afectada por factores motivacionales, perceptuales o actitudinales (Butcher, 1994).

La razón fundamental por la que hemos llevado a cabo este trabajo consiste, precisamente, en nuestra creencia en que las herramientas informáticas suponen una gran ayuda durante el complejo proceso de evaluación psicodiagnóstica (entendiendo dicho proceso como un compendio de tareas que forman parte del ámbito general de la toma de decisiones clínicas). En particular, consideramos que todas aquellas propuestas encaminadas a

solucionar problemas de diagnóstico no deben basarse sólo en aumentar el número y la calidad de los instrumentos de recogida de información, sino en la elaboración de estrategias adecuadas para el procesamiento y la combinación de los datos. En este sentido, el uso de los sistemas de diagnóstico presentados en este capítulo, desde los más convencionales (adaptación de pruebas en formato tradicional) a los más sofisticados (sistemas expertos o redes neuronales), conlleva notables ventajas que se pueden traducir en un incremento en la calidad de los diagnósticos. Por supuesto, las mejoras no se deben únicamente a la consistencia de la "máquina" para llevar a cabo tareas de manejo de la información. En parte, las ventajas evidenciadas pueden ser consecuencia de la precisión y completud de la información que se requiere para el adecuado funcionamiento de los programas (mejora de los procesos de medición y recogida de datos), o de la confrontación del "output" ofrecido por el ordenador con la propia opinión de los profesionales (si se evidencia discordancia, se emprenderán las acciones pertinentes para una nueva confirmación).

Asimismo, en relación a lo expuesto en este apartado, también resulta interesante la postura reconciliadora de autores como Hammond et al. (1987) y Liang, Chandler y Han (1990), quienes proponen distintas estrategias para obtener juicios que combinen la intuición con el razonamiento cuasi-racional y analítico (todo depende de lo suficiente y/o necesaria que sea cada una de dichas fuentes de juicio). Según la propuesta de Hammond et al. (1987), existe un continuo cognitivo que va de la intuición al análisis; la importancia de este esquema reside en la posibilidad de estudiar cómo se puede optimizar el proceso de toma de decisiones minimizando los errores que pueden estar presentes en ambas formas de razonamiento. De otro lado, Liang et al. (1990) proponen distintas estrategias para combinar el uso de procedimientos actuariales y el razonamiento inductivo desarrollando nuevos algoritmos para la resolución de problemas clínicos.