

LEADERSHIP

What is leadership? How can we tell who is the leader of a group? According to Brown (2000, p. 91), “What really characterizes leaders is that they can influence others in the group more than they themselves are influenced.”

LEADERS’ CHARACTERISTICS

It is often assumed that leaders differ from followers in intelligence and personality. For example, leaders may tend to be cleverer, more extraverted, and more dominant individuals than followers. This theory sounds plausible, but has attracted much criticism. For example, leaders seem to differ enormously in personality. It is difficult to find much similarity among George Bush in the US, Gordon Brown in the UK, and Nelson Mandela in South Africa. Most of the evidence only weakly supports the theory. Mann (1959) reviewed numerous studies. More than 70% of them showed a positive relationship between perceived leadership status and intelligence, adjustment, extraversion, dominance, masculinity, and conservatism. However, the relationships were weak: “In no case is the median correlation between an aspect of personality . . . and performance higher than 0.25, and most of the median correlations are closer to 0.15.”

The great person theory is often dismissed out of hand, but its deficiencies have been exaggerated. Lord, de Vader, and Alliger (1986) pointed out that correlations between personality measures and leadership may be low because of unreliability of measurement. They carried out a meta-analysis of the studies discussed by Mann (1959), correcting for unreliability of measurement. The correlation between intelligence and leadership perception increased from +.25 to +.52 and that between masculinity–femininity and leadership perception went up from +.15 to +.34.

So far we have focused on the personality characteristics of those *perceived* to have leadership status. A separate issue is to identify the personality characteristics of *effective* leaders in terms of group performance. Heslin (1964) reviewed the literature and found that the intelligence and adjustment of the leader correlated strongly with group performance.

There has been much controversy as to whether men or women make better leaders. Eagly, Karau, and Makhijani (1995) carried out a meta-analysis and found there were practically no overall gender differences in effectiveness. However, men performed slightly better than women in masculine leader roles (e.g., the military), whereas women performed slightly better than men in feminine leader roles (e.g., in educational or government organizations).

The great person theory doesn’t acknowledge that the qualities needed for effective leadership depend on the *situation*. For example, a more aggressive approach is needed

Do leaders share any common characteristics?



to lead an adolescent gang than a group discussing flower arranging. Hains, Hogg, and Duck (1997) studied groups of college students meeting to discuss a social issue. In each group, a member was selected at random as leader. When the members identified strongly with the group, leaders with views close to the average in the group were rated as more effective leaders than those with views differing from group members. Thus, effective leaders need to embody the group norms to be effective, which implies that the characteristics of effective leadership are much more *flexible* and situation-specific than implied by the great person theory.

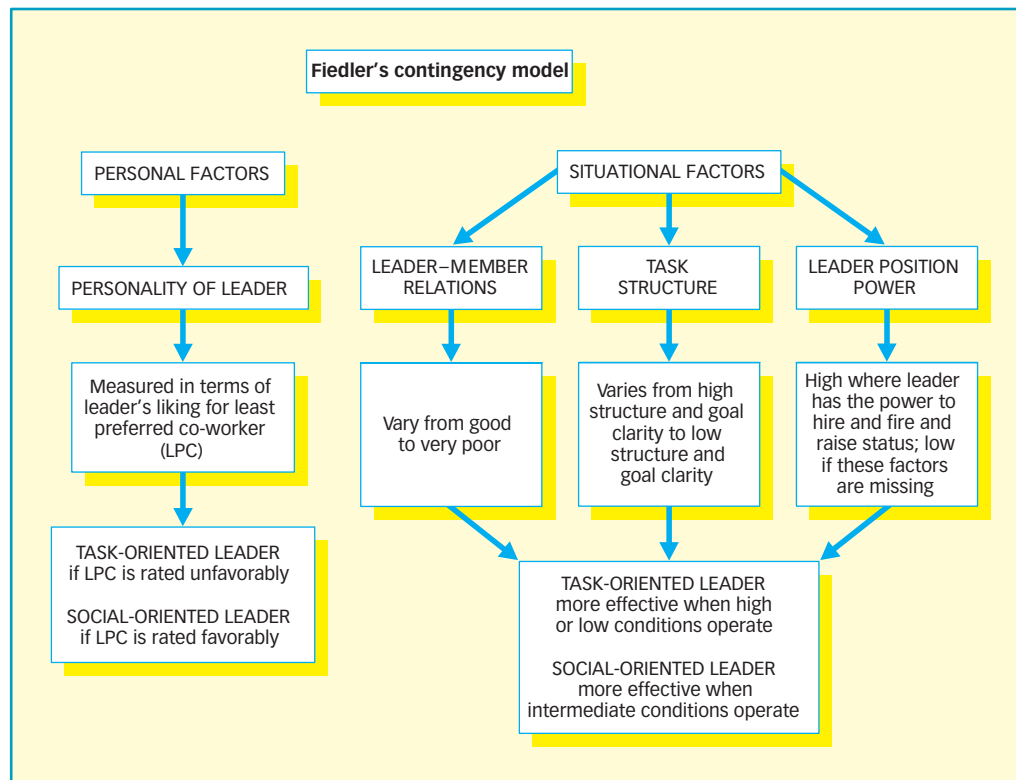
FIEDLER'S CONTINGENCY MODEL

It is important to study both the leader's personality *and* the situations in which leadership is exercised, because the effectiveness of any given leadership style is contingent on [depends on] the conditions in which the group finds itself. The most influential contingency model is the one proposed by Fiedler (1967, 1978; see the figure below).

There are four basic components in Fiedler's contingency model. One refers to the leader's personality and the others refer to features of the situation in which the leader must lead. The leader's personality is assessed on the basis of his/her liking for the least preferred co-worker. The least preferred co-worker (LPC) scale requires leaders to rate the most difficult person with whom they work. High scorers (high LPC) evaluate their least preferred co-worker relatively favorably, and are said to adopt a relationship-oriented leadership style. In contrast, low scorers (low LPC) are task-oriented.

Three situational factors jointly determine the favorableness of the situation for the leader:

1. *Leader-member relations*: The relations between the leader and the other group members can vary from very good to very poor.
2. *Task structure*: The amount of structure in the task performed by the group can vary from high structure and goal clarity to low structure and goal clarity.



3. *Position power*: The power and authority of the leadership position are high if the leader can hire and fire, raise pay or status, and has support from the organization, but are low if these factors are missing.

Any given leadership situation can be categorized as high or low on each situational factor, giving eight possible combinations or levels of situation favorableness. The most favorable situation for a leader involves good leader–member relations, high task structure, and high position power. The least favorable situation involves poor leader–member relations, a lack of task structure, and low position power for the leader. According to Fiedler, the most important situational factor is leader–member relations and the least important is position power.

We turn now to the key predictions of the model. Task-oriented leaders (low LPC) will be more effective than relationship-oriented leaders (high LPC) when the situation is very favorable or very unfavorable, but less effective when the situation is of intermediate favorableness. Why did Fiedler make these predictions? When the situation is very favorable, the leader doesn't need to be unduly concerned about relationship issues because group morale is very high. When the situation is very unfavorable, feelings within the group probably can't be improved much, and so it is best for the leader to focus on the task in hand. When the situation is of intermediate favorableness, a relationship-oriented leader can improve group morale.

Findings

Fiedler and Potter (1983) summarized the findings from over 100 studies on leadership effectiveness based on group performance. Task-oriented (low LPC) leaders were generally more successful than relationship-oriented (high LPC) leaders when the situational favorableness was low or high. However, the opposite was the case when the level of situational favorableness was moderate.

Schriesheim, Tepper, and Tetrault (1994) carried out a meta-analysis based on data from 1282 groups. Their findings were in broad agreement with those of Fiedler and Potter (1983). However, small differences in the favorableness of the situation sometimes produced large differences in the effectiveness of task-oriented or relationship-oriented leaders, which is contrary to the theory.

Fiedler argued that the most important situational factor was leader–member relations and the least important was position power. Singh, Bohra, and Dalal (1979) asked people to rate several situations for their favorableness to the leader. The relative importance of the three situational factors varied across situations, indicating that Fiedler's approach is too rigid. Contrary to the theory, position power was important in each of four studies, and it was the most influential situational factor in two of them.

Miller, Butler, and Cosentino (2004) extended Fiedler's model by relating *followers'* personality and situational favorability to their performance. The participants were junior personnel serving with the US Army in Europe. As might be expected on Fiedler's approach, relationship-oriented followers performed better than task-oriented ones in moderately favorable situations, whereas the opposite was the case with highly unfavorable situations. However, the additional finding that relationship-oriented followers performed better than task-oriented ones in very favorable situations is inconsistent with Fiedler's general approach.

Evaluation

- + The effectiveness of leaders depends on the relationship between their personal characteristics and the particular situation in which they find themselves.
- + Task-oriented leaders are most effective when the situation is very favorable or unfavorable, whereas relationship-oriented leaders are most effective when the

situation is of intermediate favorableness. Similar findings have been obtained when relationship-oriented followers are compared against task-oriented followers.

- It is oversimplified to argue that leaders are either task- or relationship-oriented. Those elected by their groups as leaders are often task- *and* relationship-oriented (Sorrentino & Field, 1986).
- Fiedler's notion that an individual's leadership style is invariant across situations is too static given that many individuals change their style to fit changing situations (Hogg & Vaughan, 2005).
- According to the contingency model, leader-member relations are solely a function of the situation. However, characteristics of the leader (e.g., warmth; agreeableness) also influence his/her relations with others within the group.
- The model tells us little about the dynamic processes occurring over time. For example, what factors determine the rise and fall of leaders?

TRANSACTIONAL VS. TRANSFORMATIONAL LEADERSHIP

Several theorists have argued that the most effective leaders are those variously described as charismatic, inspirational, or transformational. For example, Bass (1985) distinguished between transactional and transformational leadership. **Transactional leadership** involves various exchanges or bargains between leaders and followers. In contrast, **transformational leadership** involves providing one's followers with inspiration, and persuading them to rise above their own self-interests to achieve the leader's vision. Bass assumed that transactional leadership can be effective, but that transformational leaders are generally even more effective, especially when the situation is difficult and rapid changes are required. According to Bass (1985, p. 154), "Transformational leadership is more likely to reflect social values and to emerge in times of distress and change while transactional leadership is more likely to be observed in a well-ordered society."

Avolio, Bass, and Jung (1999) clarified what is involved in transactional and transformational leadership. They reported findings based on 3786 participants who described their leader using the Multifactor Leadership Questionnaire. Six factors were identified:

1. *Charismatic/inspirational*: Leader provides followers with a clear sense of purpose, and persuades them to identify with him/her.
2. *Intellectual stimulation*: Leader encourages followers to question conventional ways of solving problems.
3. *Individualized consideration*: Leader tries to understand the needs of his/her followers, and to get them to develop their potential.
4. *Contingent reward*: Leader makes it clear what he/she expects from followers, and how they will be rewarded if successful.
5. *Active management-by-exception*: Leader monitors the performance of followers, and helps to correct problems.
6. *Passive-avoidant leadership*: Leader becomes involved only when problems become serious.

An additional factor analysis of the data produced two correlated factors of transformational and transactional leadership. The factor of transformational leadership was based mainly on the charismatic/inspirational and

Key Terms

Transactional leadership:

a form of leadership involving exchanges or bargains between the leader and his/her followers.

Transformational leadership:

a form of leadership based on inspiration or charisma.

intellectual stimulation factors, whereas the transactional factor was based on individualized consideration and contingent reward.

Findings

Most studies have found that transactional and transformational leadership styles are effective. Howell and Hall-Merenda (1999, p. 681) discussed several studies on transactional leadership: “The majority of research findings suggest that contingent reward leadership [an important aspect of transactional leadership] has a positive effect on individual follower performance.” Yammarino, Spangler, and Bass (1993) carried out a longitudinal study of 193 graduates from the US Naval Academy. Measures of transformational leadership obtained at the start of the study predicted later performance appraisal.

Lowe, Kroeck, and Sivasubramiam (1996) carried out a meta-analysis. Transformational leadership was generally superior to transactional leadership in producing effective performance in work groups. This superiority of transformational leadership was obtained with groups as diverse as student leaders in laboratory studies, nursing supervisors, and German bank managers.

Kirkpatrick and Locke (1996) pointed out that most studies are correlational. Transformational leadership may correlate with follower performance because of the impact of the leader on his/her followers. However, followers may use their knowledge of the leader’s success when deciding whether he/she is transformational. Kirkpatrick and Locke distinguished between these possibilities in a study in which trained actors pretended to be transformational or nontransformational leaders. Group performance on a simulated production task was higher when the leader communicated a vision (e.g., instilling confidence in the followers; raising expectations for high performance) and when he/she provided suggestions for implementing the vision. Thus, aspects of transformational leadership can have a direct influence on group performance. However, leaders adopting a charismatic communication style (e.g., sounding dynamic and confident; making direct eye contact; animated facial expression) were no more effective than those adopting a neutral or noncharismatic communication style.

Bass (1985) argued that transformational leadership is especially likely to be effective when the situation is stressful and uncertain. This prediction has rarely been tested because most studies have focused on relatively stable conditions. However, it was tested by Bass, Avolio, Jung, and Berson (2003), who studied US Army platoons performing combat simulation exercises. In fact, platoon performance was predicted equally well by platoon leaders’ ratings on transformational leadership and on transactional leadership, which is inconsistent with the original prediction. The beneficial effects of both types of leadership on performance depended in part on increased cohesion within the platoon.

Eagly, Johanssen-Schmidt, and van Ergen (2003) carried out a meta-analysis of 45 studies comparing the leadership styles of men and women. The gender differences were small. However, female leaders tended to be more transformational than male ones, and also made more use of the contingent reward component of transactional leadership. Male leaders were more likely than female ones to use the active management-by-exception and passive-avoidant components of transactional leadership. These findings suggest the value of female leadership: “All of the aspects of leadership style on which women exceeded men relate positively to leaders’ effectiveness whereas all of the aspects on which men exceeded women have negative or null relations to effectiveness” (Eagly et al., 2003, p. 569).

Some transformational leaders have misused their skills to produce tragic consequences. For example, Marshall Applewhite was a transformational leader who claimed that his Heaven’s Gate group had reached a new stage of evolution in which they had no further need for their bodies. After discarding their bodies, they would go off in a spaceship traveling with the Hale-Bopp comet. On the basis of these claims, Marshall Applewhite persuaded nearly 40 members of his group to commit mass suicide in March 1997.

Case Study: *The Heaven's Gate Mass Suicide*

The daily papers from March 27, 1997 were full of news that 39 people had committed suicide in a hilltop mansion in Rancho Santa Fe, California. As the story broke, it became apparent that the victims were members of a cult that called itself "Heaven's Gate." The Heaven's Gate cult emerged in the 1970s and was led by Marshall Applewhite and Bonnie Nettles. They were self-described "space age shepherds" who intended to lead a flock of humans to a higher level of existence.

Through the teachings of their charismatic leaders, who claimed to be extraterrestrial representatives of the "Kingdom Level Above Human," the cult members believed their bodies were mere vessels. By renouncing sex, drugs, alcohol, their birth names, and all relationships with family and friends, disciples prepared to ascend to space, shedding their "containers," or bodies, and entering God's Kingdom. The cult members were led to believe that the appearance of comet Hale Bopp was a sign to move on to a more pure existence in outer space.

Investigations revealed that the mass suicide appeared to be a carefully orchestrated event. It took place over 3 days and involved three groups, proceeding in a calm, ritualistic fashion. Some members apparently assisted others and then went on to take their own dose of a fatal mixture. Lying on

cots or mattresses with their arms at their sides, the victims each carried identification. Each of the members of the organization gave a brief videotaped statement prior to their death. The essence of the statements was that they believed they were going to a better place.

Three things seem to be essential to the concept of a cult. Members think in terms of "us" and "them," with a total alienation from anyone perceived as "them." Intense, though often subtle, indoctrination techniques are used to recruit and hold members. The third ingredient is the presence of a charismatic cult leader who makes people want to follow his or her beliefs. Cultism usually involves some sort of belief that everything outside the cult is evil and threatening; inside the cult is the safe and special path to salvation through the cult leader and his or her teachings.

The cult leader must be extremely attractive to those who convert. He or she must satisfy the fundamental need to have someone to trust, depend on, and believe in totally. Charismatic leaders like Applewhite and Nettles gave purpose and meaning to the lives of their followers. Unquestioning devotion caused 38 Heaven's Gate cult members to voluntarily commit suicide. Marshall Applewhite was the 39th person to die in the mass suicide.

Evaluation

- + Transactional and transformational leadership are both effective, with transformational leadership generally being more effective.
- + Theories of transformational or charismatic leadership have focused on factors (e.g., vision for the future; instilling confidence in followers) ignored in other theories of leadership.
- + The theoretical approach assumes that leadership is a dynamic process that changes over time, which is an advance on Fiedler's approach.
- Some theoretical approaches focusing on transformational or charismatic leadership exaggerate the impact of the leader on the followers and minimize the impact of the followers on the leader.
- Too much transformational leadership "may be dysfunctional, because it imbues the leader with excessive power and fragments the group through continual change. The limits of transformational leadership are not specified" (Hogg & Vaughan, 2005, p. 330).
- Transformational leaders can be dangerous if the achievement of their goals involves the destruction of group members or some other group.

COLLECTIVE BEHAVIOR

Individuals often behave differently when in a crowd than when on their own or with a few friends. For example, lynch mobs in the southern parts of the United States murdered about 2000 people (mostly blacks) during the first half of the twentieth century. Those involved in those atrocities would not have behaved in that way if they hadn't been part of a highly emotional crowd.

Le Bon (1895) was a French journalist who put forward a famous theory of crowd behavior. According to him, a man forming part of a crowd:

descends several rungs in the ladder of civilization, he is . . . a creature acting by instinct . . . [He can be] induced to commit acts contrary to his most obvious interest and best known habits. An individual in a crowd is a grain of sand amid other grains of sand.

Le Bon referred to the “law of mental unity” driving a crowd to behave like a mob. He also used the term *social contagion* to describe how irrational and violent feelings and behavior can spread rapidly through a crowd.

Do you think Le Bon was basically correct? If you do, part of the reason is that the media focuses on crowds behaving badly and provides little coverage of well-behaved, dignified crowds. As we will see, Le Bon grossly exaggerated the mindlessness of crowds, and crowd behavior is much more diverse than he imagined.

CROWD BEHAVIOR

Reicher (1984) studied a civil disturbance in the St. Pauls area of Bristol in England involving the police and the mainly black community. There was considerable violence, with many people being seriously injured and several police cars being destroyed. However, the crowd’s behavior was much more controlled than might have been thought. The crowd displayed violence toward the police and symbols of the state (e.g., banks), but didn’t attack or destroy local shops and houses. Moreover, the crowd’s actions were confined to a small area lying at the heart of the community. If the crowd members had simply wanted to behave violently, then the violence would have spread into the surrounding areas. Finally, those involved denied they had lost their identities during the riots. The opposite was closer to the truth, because they experienced an increased sense of pride in their community.

How can we explain these unexpected findings? According to Reicher (1984), individuals in a crowd attend less than usual to themselves, focusing instead on the situation and the other crowd members to provide them with cues as to how to behave. This makes them responsive to group norms or standards. These group norms sometimes endorse taking aggressive action, but very often endorse responsible behavior. This theory was subsequently developed into the social identity model of deindividuation effects by Reicher, Spears, and Postmes (1995) (discussed later).

Marsh, Rosser, and Harré (1978) also found that crowds share a social purpose. They analyzed the behavior of soccer fans, discovering they had long-lasting social structures and patterns of behavior (e.g., ritualized aggression). Those fans showing the most ability to follow the rules and norms were very highly regarded and influential members of their groups.

The stereotype of soccer fans is that they form themselves into highly aggressive groups. Marsh et al. (1978) found that unrestrained fighting between rival fans happened very rarely. For example, soccer fans supporting the home team regarded it as their right to chase fans of the away team from the ground after the match, but the rival sets of fans usually kept their distance. Soccer fans often use violent language and make threatening gestures, but these activities rarely turn into actual fighting.

Waddington, Jones, and Critcher (1987) argued that most crowd violence depends on the *context* in which the crowd finds itself rather than on the characteristics of the individuals within the crowd. They compared two public rallies held during the coal miners’ strike in Britain in 1984, only one of which led to violence. In contrast to the peaceful rally, the violent one was controlled by the police rather than by the rally organizers. The violent rally hadn’t been



Hundreds of thousands of fans attend soccer matches each week but only a tiny percentage show hooligan behavior.

planned carefully with the police, and insufficient thought had been given to preventing large numbers of people being forced into a small area.

DEINDIVIDUATION

Le Bon (1895) argued that the anonymity of individuals in a crowd or mob can remove normal social constraints and so lead to violence. In similar fashion, Zimbardo (1970) and Diener (1980) argued for the importance of **deindividuation**, which is the loss of a sense of personal identity occurring in crowds. It is most likely to occur in conditions of high arousal, anonymity, and diffused responsibility (i.e., responsibility for what happens is spread among the members of the crowd).

According to Diener (1980), deindividuation is produced through decreased self-awareness and has the following effects:

- Poor monitoring of one's own behavior.
- Reduced concern to have social approval of one's behavior.
- Reduced capacity to think rationally.

Key Term

Deindividuation:

the loss of a sense of personal identity that occurs when individuals find themselves in a crowd.

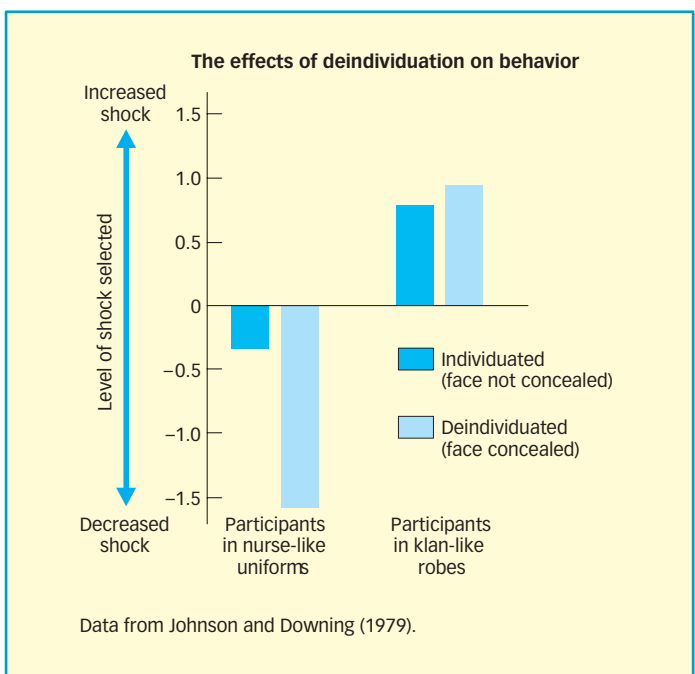


Left: Uniforms, such as those worn by nurses, increase an individual's sense of anonymity and make it more likely that they will conform to the role associated with the uniform. Right: Johnson and Downing (1979) pointed out the similarity between the clothes of Zimbardo's deindividuated participants and those of the Ku Klux Klan.

Zimbardo (1970) reported a study on deindividuation. Female participants were told to give electric shocks to other women in a Milgram-type study (see earlier in chapter). Deindividuation was produced in half the participants by having them wear laboratory coats and hoods covering their faces. In addition, the experimenter addressed them as a group rather than as individuals. The intensity of shocks given by the deindividuated participants was *twice* as great as that of participants who wore their own clothes and were treated as individuals.

Johnson and Downing (1979) pointed out that the clothing worn by the deindividuated participants in Zimbardo's (1970) study resembled that worn by the Ku Klux Klan (a secret organization that carried out many violent acts against American blacks). Deindividuated individuals dressed as nurses actually gave *fewer* electric shocks than those wearing their own clothes. Thus, deindividuation can have desirable rather than undesirable effects on behavior (see figure on the right).

Mann (1981) analyzed newspaper accounts of crowds watching someone threatening to commit suicide by



jumping from a bridge or building. Crowds often encouraged the potential suicide to jump. This aggressive crowd behavior was much more likely when those in the crowd were fairly anonymous (and thus deindividuated) because the crowd was large or the incident took place after dark. Similar findings were reported by Silke (2003), who studied 500 violent attacks carried out in Northern Ireland, 206 of them by offenders wearing disguises to mask their identities. Disguised offenders (who can be regarded as deindividuated) inflicted more serious physical injuries on their victims, attacked more people, and were more likely to threaten victims after their initial attack.

Evaluation

- + Deindividuation can contribute toward groups and crowds behaving in an anti-social or aggressive way.
- + Anonymity (a key part of deindividuation) often leads groups to behave badly because it reduces the chances that individuals will be punished for behaving anti-socially or illegally.
- Deindividuation doesn't always lead groups and crowds to behave badly; indeed, it can have the opposite effect (e.g., Johnson & Downing, 1979).
- Other theoretical approaches (see below) provide superior explanations of group behavior.

EMERGENT-NORM THEORY

According to Turner and Killian's (1972) emergent-norm theory, two factors must be present for individuals in crowds to behave aggressively:

1. The crowd should develop a group norm or standard endorsing the use of aggression.
2. Individuals in the crowd should be identifiable, because this increases the social pressures on them not to deviate from the group norm.

When both factors are present, crowd members should conform to the new or emergent norm.

We can see what is involved by considering a confrontation between a group of demonstrators and the police. A new norm that people should defend themselves against the police may emerge, which may lead to stone throwing. Individuals in the crowd who can be identified by others feel strong pressures to conform to the stone-throwing behavior.

Mann, Newton, and Innes (1982) compared the deindividuation and emergent-norm approaches. Participants observed two people having a discussion, and they could react by pressing buttons to provide crowd noise. The participants were anonymous or identifiable, and they were given fake information indicating the group norm was aggressive (loud noise) or nonaggressive (soft noise). According to deindividuation theory, anonymous individuals should have behaved more aggressively than identifiable ones, which is what Mann et al. found. However, there was some support for emergent-norm theory, because participants were more aggressive when told there was an aggressive group norm. However, that theory's prediction that the level of aggression should be greatest among identifiable participants exposed to the aggressive group norm was *not* supported. Overall the findings supported deindividuation theory more than emergent-norm theory. A problem for emergent-norm theory is that it is unclear *how* group norms are supposed to emerge.

SOCIAL IDENTITY MODEL OF DEINDIVIDUATION EFFECTS

Reicher et al. (1995) and Postmes and Spears (1998) put forward a social identity model of deindividuation effects. According to deindividuation theory, deindividuated individuals become uninhibited and freed from social constraints. According to Reicher et al.'s model, precisely the *opposite* is the case—when individuals in a group become deindividuated, their behavior is strongly influenced by the prevailing group norms.

Postmes and Spears (1998) emphasized three main assumptions of the social identity model:

1. Deindividuation leads *not* to a loss of self but only to a decreased focus on personal identity.
2. Deindividuation increases responsiveness to (or conformity with) *specific* situational group norms (i.e., what most people would regard as appropriate behavior in any given situation). This can produce very restrained behavior or aggressive behavior.
3. Deindividuation is neutral with respect to *general* social norms (standards of behavior not taking account of the particular context).

A key prediction from this theory is that deindividuation may or may not lead to anti-social behavior by group members, depending on the prevailing situational group norms. For example, consider the death of Princess Diana on August 31, 1997. This tragic event led large crowds of tearful people to show their sadness and sense of loss (in line with situational group norms) but did *not* lead to any breaking of general social norms.

Postmes and Spears (1998) carried out a meta-analysis of 60 studies on group and crowd behavior, and found only modest support for deindividuation theory. Manipulations designed to produce deindividuation (e.g., anonymity; large groups) were associated with antinormative behavior (behavior breaking general social norms). However, the average correlation or association between deindividuation manipulations and antinormative behavior was only +.09, so there was only a slight tendency for deindividuation to lead to behavior opposed to general social norms.

Other analyses by Postmes and Spears (1998) undermined deindividuation theory even further. It is assumed within that theory that manipulations of anonymity lead to antisocial behavior because they reduce the individual's self-awareness. However, the evidence provided no support for this viewpoint. Second, manipulations designed to produce deindividuation *increased* adherence to situational group norms, whereas deindividuation theory predicts a *decrease*.

The above findings can readily be explained by the social identity model of deindividuation effects. According to that model, individuals in crowds typically adopt the social identity of the crowd and their behavior is determined by situational group norms. That is precisely what was reported by Postmes and Spears (1998, p. 253):

The most striking result was that the deindividuation conditions of anonymity, larger groups, and reductions in self-awareness fostered adherence to situational norms. Thus, the factors that social psychologists have identified as playing a crucial role in the formation of collective behavior appeared to lead to a specific form of social regulation rather than its breakdown.

According to the social identity model, anonymity of individuals to the outgroup should *increase* adherence to group norms. In contrast, anonymity of individuals to the ingroup should often *reduce* adherence to group norms. The first prediction has been supported in several studies (e.g., Reicher et al., 1995). The second prediction was tested by Reicher, Levine, and Gordijn (1998). First-year psychology students responded to questionnaire statements about lying (e.g., "It is fine to give false excuses if one didn't prepare for a seminar") having been told they would discuss the various issues with a member of the academic staff afterwards. Some students were visible to the other students, whereas others sat in individual booths. As predicted, students anonymous to their fellow ingroup members were less likely to agree that lying was acceptable than those visible to other students. Visibility increased the students' perception of the power of their ingroup, and made them more willing to endorse statements contrary to the beliefs of the outgroup (academic staff).



What norms appeared as the public waited to pay their respects to Diana, Princess of Wales?

Evaluation

- + Deindividuation is important in increasing conformity behavior by crowds, and its effects can be positive (e.g., after the death of Princess Diana) as well as negative.
- + The social identity model accounts for most findings, including the controlled aggression found by Reicher (1984).
- + The theoretical assumption that deindividuation increases adherence to group norms has received much support.
- The effects of anonymity on behavior depend on how anonymity influences power relations between groups.
- Members of a large group may experience exhilaration or great excitement, but the model has little to say about such emotional states.
- It is difficult to measure key concepts such as personal identity and social identity.

Chapter Summary

Obedience to authority

- Milgram found that about 65% of people were prepared to administer potentially lethal electric shocks when an authority figure (the experimenter) ordered them to do so.
- Milgram discovered that obedience to authority was less when the obviousness of the other person's plight was increased or the authority and influence of the experimenter were reduced.
- Similar findings to those of Milgram in the United States have been found in many other countries, and obedience to authority has been shown in real-life situations.
- Milgram argued that participants were put into an "agentic" state in which they became the passive instruments of the authority figure. This seems unlikely, because most of the obedient participants experienced a strong conflict between the experimenter's demands and their own conscience.

Conformity

- Asch found that pressures to conform led many people to give the same wrong answers as other group members even though the correct answer was obvious; this is mainly a result of normative influence.
- Conformity in the Asch situation is much less when one other person in the group gives the correct answer or the other group members belong to an outgroup.
- Conformity is significantly greater in collectivistic cultures than in individualistic ones.
- According to Moscovici, minorities influence the majority through conversion (more private than public influence), whereas majorities influence minorities through compliance (more public than private influence).
- Minority influences are found mainly when the minority is perceived as part of the ingroup rather than the outgroup.
- Moscovici exaggerated the differences between the ways in which majorities and minorities exert influence.

Basic group characteristics

- The association between group cohesiveness and performance is fairly weak. There are generally stronger effects of performance on cohesiveness than of cohesiveness on performance.
- Social norms provide guidance concerning what behavior by group members is appropriate. They also help to maintain group identity, make it easier for the group to attain its goals, and promote cooperation.
- Groups (and their members) go through five phases: investigation; socialization; maintenance; resocialization; and remembrance.

- Group polarization occurs frequently in groups. It depends on persuasive arguments, social comparison, and self-categorization.
- Groupthink is most likely to occur when a group has a strong leader and there are pressures toward conformity. Political pressures help to explain many well-known examples of groupthink.

Leadership

- According to the great person theory, leaders are cleverer, more extraverted, and more dominant than their followers. There is some support for this theory, but it ignores important situational factors.
- According to Fiedler's contingency model, task-oriented leaders are more effective than relationship-oriented leaders when the situation is very favorable or unfavorable but less effective when the situation is of intermediate favorableness.
- The contingency model assumes mistakenly that all leaders are either task- or relationship-oriented but not both.
- According to Bass, transformational leadership is generally more effective than transactional leadership.
- Theories focusing on transformational or charismatic leadership minimize the impact of followers on the leader.

Collective behavior

- According to Le Bon, crowds behave in irrational and violent ways as a result of social contagion. Much of the evidence fails to support this viewpoint.
- Several theorists have argued that deindividuation is partly responsible for crowds behaving in antisocial ways. However, it doesn't always have that effect.
- According to emergent-norm theory, a crowd will behave violently when there is a group norm endorsing the use of aggression and individuals in the group are identifiable. There is mixed evidence on this theory.
- According to the social identity model, deindividuation increases conformity with situational group norms (which may endorse violent or restrained behavior). This approach has proved successful, but doesn't fully explain the strong emotional states often found in crowd members.

Further Reading

- Baron, R.S., & Kerr, N. (2003). *Group process, group decision, group action* (2nd ed.). Buckingham, UK: Open University Press. Several of the topics discussed in this chapter are dealt with at greater length in this textbook.
- Blass, T. (2000). *Obedience to authority—Current perspectives on the Milgram paradigm*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc. This interesting edited book is devoted to assessing the relevance of Milgram's ideas and research to real-world events in which enormous suffering was caused by excessive obedience to authority.
- Cialdini, R.B., & Goldstein, N.J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55, 591–621. This paper provides a good overview of current theory and research in areas such as conformity and obedience to authority.
- Goethals, G.R., & Sorenson, G. (Eds.). (2004). *Encyclopedia of leadership*. Thousand Oaks, CA: Sage. There are numerous interesting short contributions on psychological approaches to leadership.
- Hogg, M.A., & Vaughan, G.M. (2005). *Social psychology* (4th ed.). Harlow, UK: Prentice Hall. Several chapters in this textbook (especially 7, 8, and 9) contain comprehensive coverage of the topics discussed in this chapter.
- Kruglanski, A.W., & Higgins, E.T. (2003). *Social psychology: A general reader*. Hove, UK: Psychology Press. This book contains important articles on most of the topics discussed in this chapter.

Chapter 20

Contents

Social identity	485
Stereotypes	487
Prejudice and discrimination	493
Reducing prejudice and discrimination	499

Intergroup processes

Everyday life is full of examples of interactions between groups. Some of these interactions are of major political or historical importance, as is the case with the numerous wars fought between different groups or countries. Other interactions are on a more minor scale (e.g., competitive team sports; work discussions between different groups). If we define “group” broadly, then we encounter members of other groups (and those we perceive to be members of our own group) nearly every day of our lives. Thus, the study of intergroup processes and behavior is of great importance.

What exactly do we mean by intergroup behavior? According to Hogg and Vaughan (2005, p. 392), “Any perception, cognition or behavior that is influenced by people’s recognition that they and others are members of distinct social groups is intergroup behavior.” Social psychologists have devoted considerable attention to the problems that can develop between groups. This is reflected in our coverage of intergroup processes, which includes sections on stereotypes and on prejudice and discrimination. Finally, we consider how prejudice and discrimination can be reduced or even eliminated.

SOCIAL IDENTITY

Suppose someone asked you to describe your best friend in detail. Your description would certainly refer to their personal qualities (e.g., their personality). However, it would probably also include some indication of the groups to which they belong (e.g., student at college; member of the hockey team). According to social identity theory (Tajfel, 1978, 1981), the groups to which we belong form an important part of our self-concept. The term **social identity** is used to refer to an individual’s sense of himself/herself based on group membership. More specifically, “We have as many social identities as there are groups that we feel we belong to” (Hogg & Vaughan, 2005, p. 127).

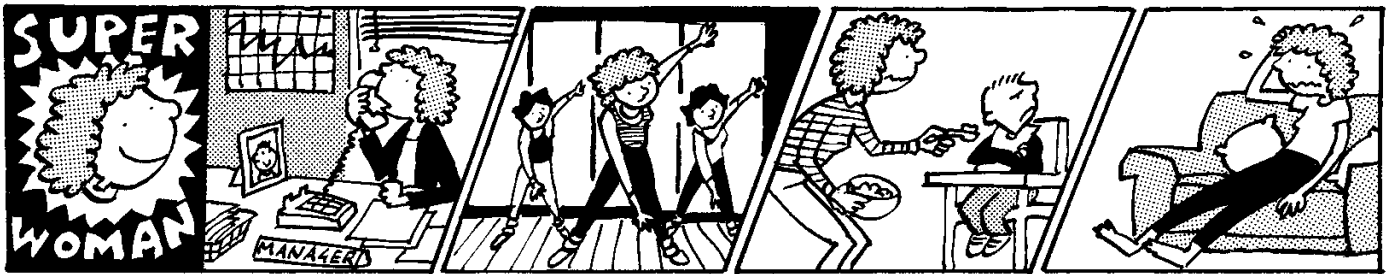
Why is it important for us to possess social identities? According to social identity theory, having a positive social identity makes us feel good about ourselves and enhances our self-esteem. One way we can achieve a positive social identity is to draw favorable comparisons between a group to which we belong (the ingroup) and some relevant outgroup (**ingroup bias** or favoritism). Drawing favorable comparisons between our ingroup and outgroups is important to our social identities. As Tajfel (1979, p. 188) expressed it, “We are what we are because *they* are not what we are.”

The good news about social identity is that it increases our self-esteem and gives us a sense of belonging. However, there is also bad news. If we boost our self-esteem by comparing our ingroup favorably to various outgroups, this may lead us to display prejudice and discrimination toward members of those outgroups. The downside of social identity is discussed later in the chapter.

Key Terms

Social identity: the part of the **self-concept** depending on the various groups to which we belong.

Ingroup bias: the tendency to favor one’s ingroup over one or more outgroups.



Findings

There is much evidence of ingroup bias or favoritism toward the ingroup at the expense of the outgroup. For example, Tajfel, Flament, Billig, and Bundy (1971) assigned 14- and 15-year-old boys to two groups. They asked individuals within the groups to assign points (which could be exchanged for money) to other participants. Nearly all the boys showed ingroup bias by awarding more points to members of their own group than the other group.

According to social identity theory, ingroup bias occurs as a *direct* result of individuals identifying themselves with the ingroup, and so should always be found. However, there is an alternative explanation based on self-interest. Perhaps individuals reward ingroup members more than outgroup members because they expect to be benefited in return by other ingroup members. Support for the notion that self-interest is involved in ingroup bias was reported by Gaertner and Insko (2000). Some participants were told they wouldn't receive any bonus money in contrast to the other ingroup members and most of the outgroup members. These participants didn't show any ingroup bias because other ingroup members couldn't benefit them in return.

Most laboratory studies on ingroup bias have involved the allocation of positive outcomes (e.g., points; money). According to social identity theory, individuals should also show ingroup bias or favoritism when allocating *negative* outcomes (e.g., punishments). Support for this prediction was reported by Verkuyten, Drabbles, and van den Nieuwenhuijzen (1999). Dutch participants indicated how strongly they identified themselves with the Dutch majority ingroup in the Netherlands. Individuals who identified themselves most strongly with the Dutch ingroup revealed the greatest negative emotions toward ethnic minorities. In similar fashion, Germans during the Second World War who identified most strongly with being German were more strongly opposed to the Jews than were Germans with less national ingroup identification (Goldhagen, 1996).

According to social identity theory, ingroup bias leads to increased self-esteem for its members. Rubin and Hewstone (1998) reviewed the literature, and found that 9 out of 12 studies reported supporting evidence. For example, Lemyre and Smith (1985) allowed some participants to give rewards to members of either an ingroup or an outgroup, and thus to show ingroup bias. The other participants had to give rewards either to one or two ingroups or to one or two outgroups, and so could not show ingroup bias. Those participants who could show ingroup bias had higher self-esteem than those unable to do so.

As we have seen, an important reason for ingroup bias is that it enhances an individual's self-esteem. Individuals who initially have low self-esteem should exhibit more ingroup bias than those with high self-esteem, because they have a stronger motive to enhance their self-esteem. This prediction has received very little support. Rubin and Hewstone (1998) found that only 3 out of 19 studies reported the predicted findings.



Strong negative emotions need to be created in order to make an ingroup justify its harmful behavior toward an outgroup.

Evaluation

- + Our self-concept depends importantly on the groups with which we identify.
- + Social identity theory has been applied successfully to several phenomena within social psychology, including ingroup bias, stereotyping, prejudice, and prejudice reduction (see later in the chapter).
- The prediction that individuals low in self-esteem should exhibit more ingroup bias than those high in self-esteem has not been supported.
- According to social identity theory, people very readily adopt social identities. However, there are (unknown) limits to this process. As Augoustinos and Walker (1995, p. 131) pointed out, “People do not accept any social identity thrust upon them, they actively seek, avoid, resist, dispute, and negotiate social identities.”
- Social identity theory focuses on the cognitive processes underlying group identification, and has little to say about emotional and motivational factors.



Members of the Star Wars Fan Club convention—a social identity?

STEREOTYPES

When we think about some group in society (e.g., Catholics; Italians), we often make use of stereotypes. A **stereotype** is, “a cognitive representation or impression of a social group that people form by associating particular characteristics and emotions with the group” (Smith & Mackie, 2000). Stereotypes are schemas or organized packets of knowledge relating to specific groups or individuals. For example, many people have a stereotype of the English as intelligent, tolerant, and reserved, even though they know many English people who are completely different from this stereotype! Note that this example is not typical, because most stereotypes are negative and related to prejudice.

It is often assumed that stereotypes are very inaccurate. However, there is often a grain of truth in stereotypes. For example, McCauley and Stitt (1978) asked various groups of Americans to guess the percentages of adult Americans and of black Americans who hadn’t completed high school, were born illegitimate, had been the victims of violent crime, and so on. There were differences in the guesses for most questions, thus showing the existence of stereotypes. Surprisingly, however, the participants *underestimated* the actual differences between the two groups, so their stereotypes had some basis in fact. In contrast, Terracciano et al. (2005) carried out a large-scale study on national character across 49 cultures. They found no relationship between national stereotypes (based on personality) and mean personality trait levels across those cultures, suggesting that there is no validity to stereotypes about national character.

ASSESSING STEREOTYPES

Traditionally, stereotypes were nearly always assessed by means of questionnaires. For example, McCauley and Stitt (1978) asked participants questions such as, “What percentage of people in the world generally are efficient?” and “What percentage of Germans are efficient?” The average answer to the former question was 50%, whereas it was 63% to the latter one. The stereotype of Germans was assessed by using several such pairs of questions focusing on large differences in responses depending on whether the questions referred to Germans or to people in general.

There are two major problems with most questionnaire measures. First, there is social desirability bias. Individuals having very negative stereotypes of other groups may feel it is socially desirable to pretend their stereotypes are less negative than is actually the case. Second, some aspects of an individual’s stereotypes may not be accessible to conscious

Key Term

Stereotype:

a simplified cognitive generalization or categorization (typically negative) about a group. It is often based on easily identifiable characteristics (e.g., sex; ethnicity).

awareness. Thus, individuals may lack the ability to report accurately on their stereotypes when completing a questionnaire.

Traditional questionnaire methods provide an assessment of *explicit* attitudes and stereotypes that is, those of which the individual is consciously aware. Some of the limitations with questionnaires can be overcome by assessing *implicit* attitudes and stereotypes, that is, those of which the individual is not consciously aware. The Implicit Association Test (IAT) was devised by Greenwald, McGhee, and Schwartz (1998) to assess unconscious stereotypes. We will consider the version of this test used by Cunningham, Preacher, and Banaji (2001) with white participants. On some trials, participants were presented with faces, and pressed one key for white faces and a second key for black faces. On other trials, they were presented with good (e.g., love, happy) and bad (e.g., poison, terrible) words, and pressed the same two keys to indicate whether each word was good or bad. In condition 1, white faces and good words were classified using one key and black faces and bad words were classified using the second key. In condition 2, white faces and bad words involved one key and black faces and good words the second key. Reaction times were much faster in condition 1 than in condition 2, suggesting the existence of implicit pro-white and anti-black stereotypes.

How different are explicit and implicit stereotypes? Cunningham et al. (2001) found that three different measures of implicit racial stereotypes all correlated positively with scores on a measure of explicit racial stereotypes (the Modern Racism Scale). The mean correlation was $+.35$, revealing a modest tendency for individuals having implicit racial stereotypes to have explicit ones as well. More strikingly, participants showed more evidence of prejudice on the implicit measures, indicating that the implicit measures assess prejudice *not* revealed by questionnaires.

Akrami and Ekehammar (2005) considered further the relationship between explicit and implicit stereotypes. In their first analysis, there was no relationship between explicit and implicit racial prejudice or stereotypes. However, they argued that the explicit scores of some participants were misleadingly low because they had a strong motivation to control their prejudiced reactions. When they “corrected” the explicit scores by taking account of their motivation to control prejudice, Akrami and Ekehammar found that the measures of explicit and implicit stereotypes correlated significantly with each other.

In sum, explicit and implicit measures of stereotypes typically overlap to some extent. However, implicit measures seem to assess important aspects of stereotypes that cannot be measured by explicit measures.



What characteristics do you think each of these people might possess?

WHY DO WE HAVE STEREOTYPES?

Nearly everyone possesses numerous stereotypes. This suggests that stereotypes probably fulfill one or more important functions. It has often been argued (e.g., Macrae & Bodenhausen, 2000) that stereotypes provide a simple and economical way of perceiving the world. Thus, for example, we can readily categorize someone we meet for the first time on the basis of their sex, age, clothing, and so on. However, stereotypes don't only minimize the amount of information processing we need to carry out. They also fulfill important social and motivational functions. They help us to achieve a sense of social identity by allowing us to distinguish ourselves clearly from the members of other groups (Oakes, Haslam, & Turner, 1994). These two approaches are considered in turn below.



Stereotypes can help us achieve a sense of social identity.

Cognitive approach

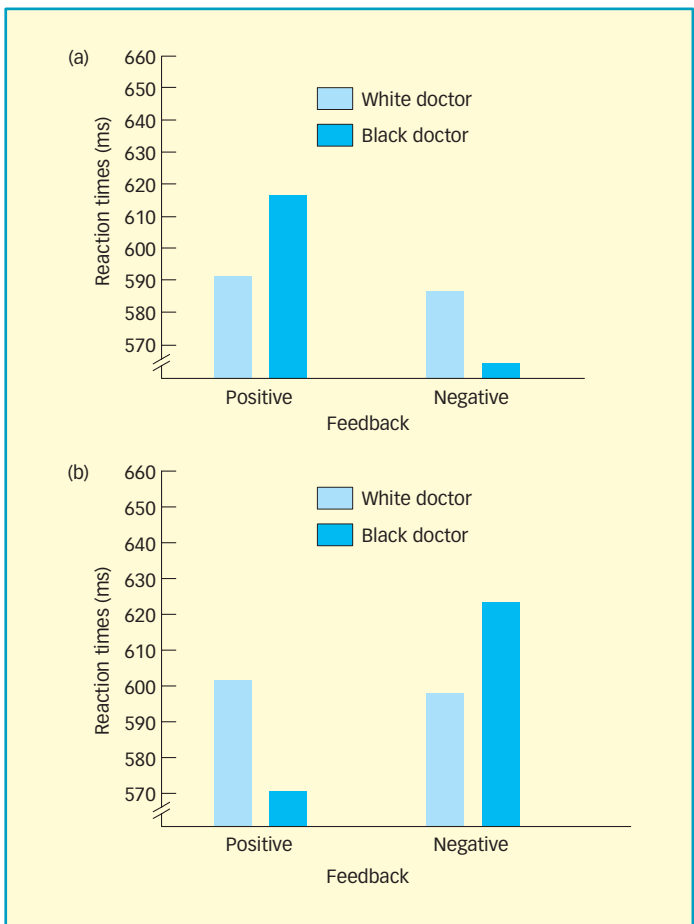
The cognitive approach is based on the assumption that stereotypes reduce the amount of processing needed when we meet or think about other people. This could happen because relevant stereotypical information is activated automatically and effortlessly whenever we encounter a member of a given group. However, matters aren't as simple as that. Most people we meet can be categorized in several different ways (e.g., female; young; French; student), and so several stereotypes should be activated. How do we decide which stereotype to focus on? According to Macrae and Bodenhausen (2000), we initially activate all relevant stereotypes in any given situation. There is then a competition for mental dominance, in which the non-dominant stereotypes are actively inhibited.

Findings

Macrae, Milne, and Bodenhausen (1994) tested the notion that stereotypes reduce cognitive processing. Participants performed two tasks at the same time. One task involved forming impressions of various imaginary people when given their names and personality traits, and the other was a comprehension task. Half the participants were told the job held by each of the imaginary people on the impression-formation task, and so could make use of stereotypical information. For example, participants told that someone was a doctor could activate stereotypical information about the kind of person having that job (e.g., intelligent; hard-working; caring).

What would we predict? If the use of stereotypes reduces the amount of processing required, participants able to use stereotypes should have performed better on both tasks than those who couldn't use stereotypes. That was exactly what Macrae et al. (1994) found.

Sinclair and Kunda (1999) focused on the issue of the factors determining which stereotype is dominant in any given situation. They argued that we activate positive stereotypes and inhibit negative ones when motivated to think well of another person. In contrast, we activate negative stereotypes and inhibit positive ones when motivated to think badly about someone else. They had a black doctor provide positive or negative feedback concerning the participant's performance on a test of



Speed of responding to black words (a) and to doctor words (b) as a function of feedback (positive or negative) received from a white doctor or a black doctor. From Sinclair and Kunda (1999). Copyright © by the American Psychological Association. Reprinted with permission.

interpersonal skills. It was assumed that participants receiving positive feedback from the black doctor would activate the doctor stereotype and inhibit the black stereotype. In contrast, those receiving negative feedback would activate the black stereotype and inhibit the doctor stereotype. The findings were precisely as predicted (see figure on the previous page). As Sinclair and Kunda (1999, p. 903) concluded, “The same individuals may be viewed through the lenses of different stereotypes by perceivers with different goals; the same Black doctor may be categorized and viewed as a doctor after delivering praise but as a Black person after delivering criticism.” There was very little evidence of changes in stereotypes when the feedback was provided by a white doctor.

Evaluation

- + Stereotypes reduce the amount of cognitive processing required.
- + Stereotypical information irrelevant in a given context is inhibited.
- As we will see shortly, stereotypes are much more *flexible* than implied by the cognitive approach.
- The nature of any given stereotype varies as a function of the particular social context in which it is used (see below).

Social approach

According to the cognitive approach, most stereotypes are relatively permanent and rigid, with the same stereotypical information being activated in different situations. A very different perspective is offered by social identity theory (discussed earlier in the chapter). According to that theory, there is considerable *variability* in the specific stereotypical information activated across situations. For example, positive stereotypical information about the French may be activated when eating in a French restaurant or visiting Paris. In contrast, negative stereotypical information about the French may be activated when listening to an arrogant French politician on television or thinking about extreme right-wing French groups.

It is often assumed that stereotypes represent irrational and invalid prejudices, and so are generally undesirable. In contrast, social identity theory claims that it is inevitable (and even desirable) that individuals categorize themselves and others on the basis of their social identities as members of groups. Thus, part of our sense of who we are stems from identifying with certain groups and not identifying with other groups.

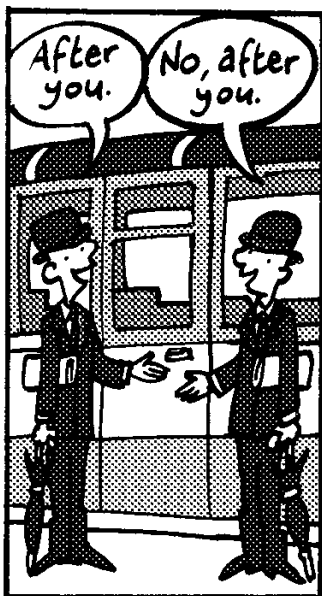
Findings

There is much evidence that the stereotypical information we focus on depends on the prevailing social context. For example, Cinnirella (1998) assigned British students to one of three tasks:

1. Provide stereotype ratings of the British only.
2. Provide stereotype ratings of the Italians only.
3. Provide stereotype ratings of the British and of the Italians.

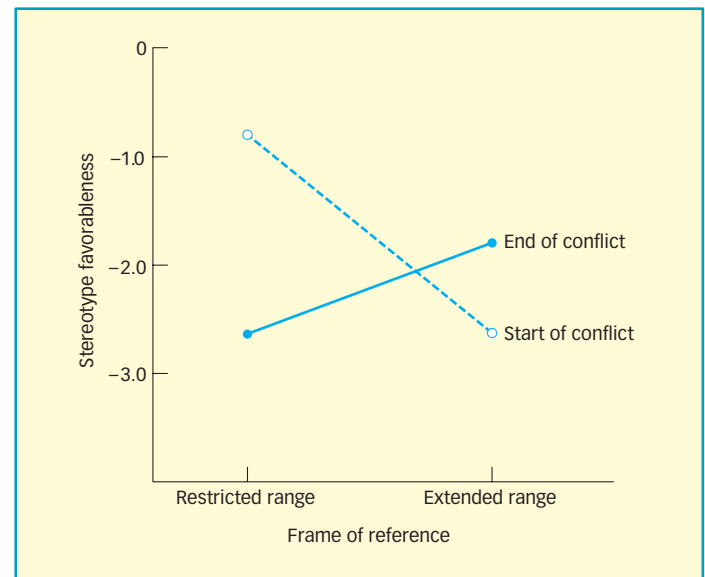
The key assumption was that participants in condition 3 would be motivated to differentiate clearly between the British and Italian stereotypes. As a result, the positive features of the British stereotype and the negative ones of the Italian stereotype should be most evident in that condition.

What did Cinnirella (1998) find? As predicted, some components of the British stereotype (industrious, reserved) were more pronounced in condition 3 than in condition 1. In addition, the Italian stereotype was more negative in condition 3 than in condition 2. The Italians were rated as less industrious, intelligent, and progressive when compared against the British than when considered on their own.



Many people have a stereotype of the English as intelligent, tolerant, and reserved.

Haslam et al. (1992) also found that stereotypes are flexible and influenced by social context. They assessed the stereotypes of Americans possessed by Australian students at the start and end of the Gulf War in 1991. The students did this either using Australia and Britain as the frame of reference (restricted range) or with Australia, Britain, Iraq, and the Soviet Union as the frame of reference (extended range). The favorability of the American stereotype changed as a function of the time of testing (start vs. end of conflict) and the frame of reference (see figure on the right). When the frame of reference included only Australia and Britain, the students (most of whom were anti-war) showed a reduction in stereotype favorableness over time. However, when the frame of reference included Iraq (a country with whom Australia was in conflict), the favorability of the American stereotype *increased* over time. This occurred because America was regarded as an ingroup when compared against Iraq. These findings show that stereotypes can vary considerably as a function of the specific context in which they are elicited.



Stereotype favorableness for Americans held by Australian students as a function of time (start vs. end of 1991 Gulf War) and frame of reference (restricted range vs. extended range). Data from Haslam et al. (1992).

Evaluation

- + Stereotypes are influenced by the immediate social context. Thus, they are flexible, and not unchanging as implied by the cognitive approach.
- + Perceiving oneself and others in terms of social identities and stereotypes is natural and inevitable.
- In spite of much variability in stereotypes across situations, it is still likely that most stereotypes possess a fairly unchanging central core of meaning as suggested by cognitive theorists.
- We often can't predict the precise nature of an individual's stereotypes in any given situation. As Turner (1999, p. 26) pointed out, "Like all perception, they [stereotypes] vary with the expectations, needs, values, and purposes of the perceiver."

HOW ARE STEREOTYPES MAINTAINED?

Various factors help to maintain stereotypes after formation. In general, information consistent with our stereotypes is attended to and stored away in memory. In contrast, information inconsistent with our stereotypes is often ignored and/or forgotten. Bodenhausen (1988) studied the negative stereotypes many Americans have about people of Spanish origin. In his first study, American participants pretended they were jurors at a trial. The defendant was described to some as Carlos Ramirez, a Spanish-sounding name. To others, he was described as Robert Johnson. The participants then read the evidence, and decided how likely it was the defendant was guilty. Those who knew him as Carlos Ramirez rated him as more guilty than those who knew him as Robert Johnson. Thus, stereotypes lead to biased processing of information.

In his second study, Bodenhausen (1988) found out more about the processes involved. He argued that stereotypes might lead people to *attend* only to information fitting their stereotype, or it might lead them to *distort* the information to make it support their stereotype. In order to prevent participants from attending selectively to stereotype-fitting information, Bodenhausen asked them to rate each item of evidence immediately in terms of whether it favored or did not favor the defendant. Carlos Ramirez was no

longer rated as more guilty than Robert Johnson. Thus, stereotypes make us attend to information fitting the stereotype and cause us to disregard other items of information.

It might be imagined that most people would remember information congruent or in line with their stereotypical views and forget incongruent information. In fact, matters are actually more complicated than that. Stangor and McMillan (1992) carried out meta-analyses to establish which kind of information is better remembered. Among individuals having weak or moderate stereotypes, information *incongruent* with the stereotype was generally remembered better than congruent information. However, congruent information was better remembered than incongruent information among individuals having strong stereotypes. Thus, memory processes serve to maintain stereotypes only for those already possessing strong stereotypes.

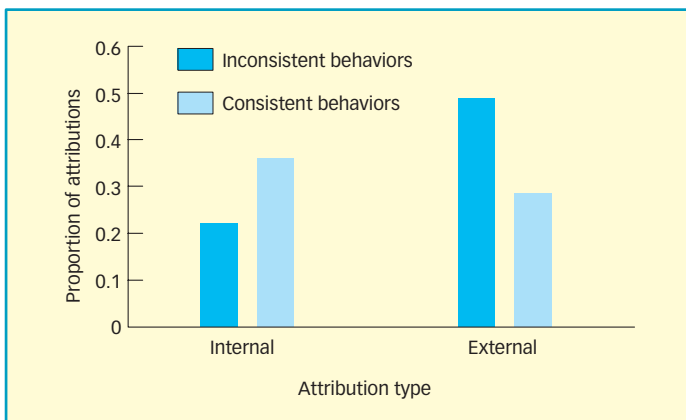
WHY IS IT HARD TO CHANGE STEREOTYPES?

It is generally remarkably difficult to produce long-lasting changes in someone's stereotypes. Why is this? Sherman, Stroessner, Conrey, and Azam (2005) argued that it is important to consider our attributions about other people's behavior. As we saw in Chapter 17, other people's behavior can be attributed to dispositional or internal causes (e.g., personality) or it can be attributed to situational causes. Sherman et al. presented participants with a range of behaviors exhibited by a gay man from Chicago called

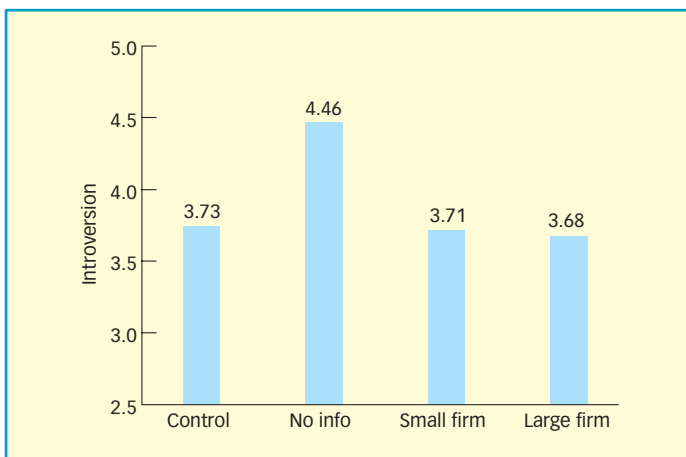
Robert. Participants prejudiced against homosexuality gave *internal* attributions to Robert's stereotype-consistent behaviour but *external* attributions to his stereotype-inconsistent behavior (see figure on the left). When someone's behavior is attributed to internal or dispositional causes, we expect that behavior to continue in the future. However, we don't expect behavior attributed to external causes to continue. Thus, even though the participants found that some of Robert's behavior was not consistent with the stereotype of gays, they nevertheless expected that in future he would mainly behave in a stereotype-consistent way.

Another popular explanation of the persistence of stereotypes is based on the subtyping model (e.g., Brewer, Dull, & Lui, 1981). According to this model, individuals violating the stereotype of their group are simply assigned to a separate subtype, and so are regarded as unrepresentative of the group. For example, suppose we have a stereotype that Germans are efficient, but we meet an inefficient German professor (they do exist!). This may lead us to conclude that members of the subtype German professors are inefficient, but all other Germans are efficient (Weber & Crocker, 1983).

Kunda and Oleson (1995) predicted that people will use almost any information about deviant (stereotype-breaking) individuals to justify subtyping them and thus regarding them as unrepresentative. Participants read a copy of an interview with an introverted lawyer called Steve. He was a deviant, because the stereotype of lawyers in the United States is that they are extraverted. In order to provide the participants with some (flimsy) grounds for subtyping Steve, some were told that he worked for a small firm or for a large firm. Other participants were given no information about the size of firm for which Steve worked, and control participants didn't read the interview. All groups provided ratings of how introverted or extraverted lawyers are at the end of the experiment.



Mean proportion of internal and external attributions for stereotype-consistent and stereotype-inconsistent behaviors for participants high in prejudice. (There were no significant effects for participants low in prejudice.) From Sherman et al. (2005). Copyright © by the American Psychological Association. Reprinted with permission.



Beliefs about lawyers' introversion without specific information (control condition) or after reading about a specific lawyer called Steve (other conditions). From Kunda and Oleson (1995). Copyright © by the American Psychological Association. Reprinted with permission.

The findings are shown in the figure on the previous page. The ratings of the control group represent the standard stereotypical view that lawyers are not introverted. Those participants given the irrelevant information that Steve worked for a small or a large firm maintained their stereotypical view that lawyers are extraverted, using that information to avoid generalizing from Steve to lawyers in general. In contrast, the no-information participants who had no grounds for regarding Steve as unrepresentative of lawyers changed their stereotypical view of lawyers in the direction of perceiving them as more introverted.

PREJUDICE AND DISCRIMINATION

Many people regard prejudice and discrimination as meaning the same thing. In fact, prejudice is an attitude or belief, whereas discrimination refers to behavior or action. According to Smith and Mackie (2000, p. 156), **prejudice** is “the positive or negative evaluation of a social group and its members.” It differs from stereotyping in that the emphasis with prejudice is more on emotional factors and less on cognitive ones. **Discrimination** involves negative actions (e.g., aggression; exclusion from society) directed at the members of some group.

It seems reasonable to assume that prejudice and discrimination would be closely related. Thus, the more negative your attitudes toward another group (prejudice) the more negative your behavior toward that group is likely to be (discrimination). In fact, Dovidio, Brigham, Johnson, and Gaertner (1996) carried out a meta-analysis. They found that prejudice correlated only +.32 with discrimination, indicating there is only a modest association between them. There are generally greater social pressures to avoid discrimination (which is readily observable by others) than to avoid prejudice (which is less obvious to other people).

Discrimination can take many forms. Allport (1954) argued that there are five stages of discrimination. In some situations (e.g., Nazi Germany), the level of discrimination increases rapidly from the early stages to the later ones. Here are Allport’s five stages:

1. *Anti-locution*: Verbal attacks are directed against some other group.
2. *Avoidance*: The other group is systematically avoided. This can involve steps to identify group members (e.g., the Star of David worn by Jews in Nazi Germany).
3. *Discrimination*: The other group is deliberately treated less well than other groups in civil rights, job opportunities, membership of clubs, and so on.
4. *Physical attack*: Members of the other group are attacked and their property is destroyed.
5. *Extermination*: There are deliberate attempts to kill all members of the other group (e.g., the gas chambers built by the Nazis to murder the Jews).

Many groups have been on the receiving end of prejudice and discrimination. However, the most vulnerable groups are those that are easily identifiable. This helps to explain why race, gender, and age are the “top three” categories used for purposes of stereotyping, prejudice, and discrimination (S. T. Fiske, 1998). We focus mainly on racism (probably the most intensively studied form of prejudice), but also consider sexism briefly.

RACISM

Racism can be defined as prejudice and discrimination against others because of their race or ethnicity. The evils of racism can be seen in the mass slaughter during the twentieth century in several countries including Germany, the former Yugoslavia, Rwanda, and South Africa. There is superficial evidence that racism is in decline in countries such as Britain and the United States. For example, S. T. Fiske (2002) argued

Key Terms

Prejudice:

an attitude or belief (usually negative) toward the members of some group on the basis of their membership of that group.

Discrimination:

negative actions or behavior directed against the members of some group.

Racism:

prejudice and/or discrimination against another group because of their race or ethnicity.



Discrimination against specific groups is sometimes aided by distinguishing visual characteristics (skin color, or style of dress, for example). Sometimes, however, minority group members are not clearly distinguishable from the majority and are forced to identify themselves. This was the case in Nazi Germany where Jews had to wear a Star of David on their clothing, making them a focus for racial hatred.



In a study by Allport and Postman (1947) participants were shown this picture. Later they were more likely to recall that the black man was holding the razor.

that only 10% of individuals in Western societies have overt and obvious racial biases. However, as many as 80% of people possess various subtle racial biases, which lead to “awkward social interactions, embarrassing slips of the tongue, unchecked assumptions, stereotypic judgments, and spontaneous neglect” (S. T. Fiske, 2002, p. 124).

Dovidio and Gaertner (1991) argued that large numbers of people exhibit aversive racism. **Aversive racism** can be defined as “attitudes toward members of a racial group that incorporate both egalitarian [belief in equality] social values and negative emotions, causing one to avoid interaction with members of the group” (Franzoi, 1996, p. 405). The notion of ambivalent racism (McConahay, 1986) is similar. Individuals with ambivalent racism experience much conflict between their beliefs in equality and sympathy for those who are oppressed and their beliefs that individuals are responsible for what happens to them. As a result, many white Americans are willing to praise successful black Americans, but have poor opinions of black Americans who appear unwilling to work hard.

Swim, Aikin, Hall, and Hunter (1995) argued there are three main ways in which modern racism (very similar to aversive racism) manifests itself. First, modern racists deny

there is prejudice and discrimination against minority groups. Second, they show annoyance and impatience at the fact that minority groups demand equal treatment with the majority group. Third, they have feelings of resentment at the prospect of minority groups receiving positive action to assist them.

We can study racism by seeing whether a given ambiguous situation is interpreted differently depending on whether the central figure is, say, white or black. This was done by Duncan (1976), who asked white American students to watch a conversation between a black man and a white man. When the white person gently shoved the black person, the action was interpreted as violent by only 13% of the participants. In contrast, 73% of the participants interpreted the same action as violent when the black man did the shoving.

Racial bias influences even basic perceptual processes. Payne (2001) presented a photograph of a male face (white or black) briefly. The photograph of an object was then presented, and participants decided rapidly whether it was a handgun or a handtool. White participants were more likely to identify a tool mistakenly as a gun when preceded by a black face than when preceded by a white face, and this seems to have happened automatically. Individuals high in explicit prejudice showed more racial bias on this task than those low in explicit prejudice. Allport (1954) discussed his famous study in the United States in which white Americans saw the drawing top left. They were asked afterwards to indicate who had the knife in his hand. Many participants misremembered (and perhaps misperceived), claiming that it was the black man.

SEXISM

Sexism involves prejudice against individuals purely on the basis of their sex. Probably the commonest sexist assumptions are that women are more caring than men, but men are more assertive and competent. These assumptions have been found in many parts of the world, including Australia, Europe, North America, and South America (Deaux, 1985).

Sexism is apparent in what is known as the “glass ceiling”: women in most Western societies are as well qualified as men, but occupy under 10% of top positions (e.g., CEO). Eagly and Karau (2002) explained this in their role congruity theory. According to this theory, women are thwarted in their attempts to achieve top positions because of two forms of prejudice:

1. Men typically evaluate women’s potential for leadership less highly than that of men. This is because there are large discrepancies between the perceived qualities needed for

Key Term

Aversive racism: a combination of beliefs in equality for all and negative emotions toward members of other races.

leadership (e.g., decisive; action-oriented; courageous) and perceptions of the female gender role (e.g., helpful; sympathetic; nurturant).

2. Leadership behavior that is decisive and action-oriented and thus conforms to expectation is regarded more favorably when shown by a man than by a woman.

Davison and Burke (2000) reviewed several studies in which half the participants received a résumé (CV) with a male name attached to it and the other half received the same résumé with a female name attached. Men were preferred to women for jobs rated as male sex-typed (i.e., involving skills typically associated with men). The relevance of this finding is that most leadership positions are male sex-typed.

Additional support for role congruity theory comes from Eagly, Makhijani, and Kinsky (1992) in a review of studies in which leadership behavior was described and ascribed to a man or a woman. In their meta-analysis of 61 studies, this behavior was rated less highly when performed by a woman. Of most importance, the tendency to devalue women's leadership behavior was much greater when the behavior was stereotypically masculine (e.g., directive or autocratic).

In sum, there is much support for role congruity theory. However, the prejudicial effects obtained in most of the studies and meta-analyses accounted for only 1–5% of the variability in the data. Thus, it is not clear whether such effects can account fully for the “glass ceiling.”

AUTHORITARIAN PERSONALITY

Adorno, Frenkel-Brunswik, Levinson, and Sanford (1950) argued that people with an authoritarian personality are most likely to be prejudiced. The **authoritarian personality** includes the following characteristics:

- Rigid beliefs in conventional values.
- General hostility toward other groups.
- Intolerance of ambiguity.
- Submissive attitude toward authority figures.

How does the authoritarian personality develop? According to Adorno et al. (1950), it has its roots in childhood experiences. Children receiving a harsh upbringing with little affection and much punishment from their parents are most likely to develop an authoritarian personality. This treatment causes the child to have much hostility toward his/her parents. However, this hostility remains unconscious because the child is unwilling to admit to it. This hostility is displaced on to nonthreatening minority groups, and appears in the form of prejudice. Thus, the hostility that harshly treated children can't express toward their parents is later redirected onto innocent groups.

Adorno et al. (1950) devised various questionnaires. One was the F (Facism)-Scale, designed to measure the attitudes of the authoritarian personality. Here is a sample item: “Most of our social problems would be solved if we could somehow get rid of the immoral, crooked, and feeble-minded people.”

Adorno et al. (1950) found that high scorers on the F-Scale were more prejudiced than low scorers. Pettigrew and Meertens (1995) reported that authoritarian individuals in France, the Netherlands, Britain, and Germany were prejudiced against a wide range of outgroups. Adorno et al. also found (as predicted by their theory) that high scorers on the F-Scale reported being treated more harshly than nonauthoritarian individuals during childhood. However, the finding of an association between certain childhood experiences and an authoritarian personality doesn't prove that the childhood experiences *caused* the authoritarian personality.

Key Term

Authoritarian personality: a type of personality characterized by rigid beliefs, hostility toward other groups, and submissive attitudes toward those in authority.



Children who receive a harsh upbringing with little parental affection are most likely to develop an authoritarian personality. The child's anger toward his/her parents is repressed only to emerge in later life in the form of prejudice.

The nine personality traits of the authoritarian personality, from Adorno et al.'s F-Scale

Traits	Description
Conventionalism	Very conventional, great dislike of change
Authoritarian–Submissive	Deferential to authority
Authoritarian–Aggressive	Very hostile to people who challenge authority
Anti-inception	Very intolerant of behavior that is “Wrong” in any way
Superstition & stereotype	Believes in fate
Power & “toughness”	Has a dominating and bullying manner
Destructiveness & cynicism	Very hostile toward anyone with whom they disagree
Projectivity	Projects own unconscious impulses onto other people
Sex	Has an exaggerated interest in sexual behavior that is not regarded as “normal”

Adorno et al.'s approach has several limitations. First, Altemeyer (1988) obtained evidence suggesting that the roots of authoritarianism lie in adolescence rather than early childhood. Adolescents whose parents are authoritarian imitate their parents' behavior, and are often rewarded for doing so.

Second, measures of the authoritarian personality assess social attitudes rather than personality. Personality changes little over time but it is easy to produce changes in authoritarianism scores. For example, Altemeyer (1988) found that showing people scenarios concerning threatening social changes produced substantial increases in authoritarianism scores.

Third, and most important, prejudice depends more on cultural norms than on individual personality. For example, Pettigrew (1958) found that the levels of authoritarianism were the same in South Africa and in the United States. However, there was much more prejudice in South Africa than the United States because of political and cultural factors.

REALISTIC GROUP CONFLICT

Sherif (1966) argued that prejudice often results from intergroup conflict (see Jackson, 1993, for a review). Each group has its own interests and goals. When two groups compete for the same goal this creates realistic conflict, which can cause the members of each group to become prejudiced against the members of the other group. That is the central assumption of realistic conflict theory. Sometimes two groups have the same interests and are pursuing the same goal. When that happens, the two groups will often cooperate with each other and there will be an absence of prejudice.



People who feel strongly about a particular cause are sometimes likely to experience violent clashes with people who do not share the same values. Here, the Argentinean Piqueteros movement, supporters of President Cristina Fernandez's government, clash in Buenos Aires with demonstrators from the opposition (farmers protesting against President Fernandez's tax hike on grain exports).

Findings

In a famous study by Sherif et al. (1961), prejudice was created between two groups of ordinary boys at a summer camp—see the Key Study on the following page.

Sherif et al.'s (1961) key findings have been replicated in various cultures. However, competition doesn't always lead to prejudice and intergroup conflict. Tyerman and Spencer (1983) argued that competition mainly has dramatic effects when those involved are initially strangers, as in the Sherif et al. study. Tyerman and Spencer observed scouts who already knew each other well as they competed in groups against each other in their annual camp. Competition didn't produce the negative effects obtained by Sherif et al.

Key Study

Sherif et al. (1961): The Robber's Cave study

In this study 22 boys who didn't initially know each other spent 2 weeks at a summer camp in the United States. There were two groups (the Eagles and the Rattlers), who were told that the group performing better in various competitions would receive a trophy, knives, and medals. This competition led to a fight breaking out between the members of the two groups, and the Rattlers' flag was burned.

There was much prejudice. Each group regarded its own members as friendly and courageous whereas the members of the other group were smart-alecks and liars. When the boys identified their friends, 93% of them were members of the same group and only 7% came from the other group.

Various attempts were made to reduce the conflict between the Eagles and the Rattlers. For example, the camp's drinking water was turned off, and the two groups had to combine forces to restore the supply. Other tasks requiring cooperation between the groups were rescuing a truck that had got stuck and pitching tents. A consequence of pursuing these common goals was that the groups became much friendlier toward each other. When asked again to identify their friends at the camp, 30% of their choices after these cooperative activities were members of the other group.

Discussion points

1. Why has the study by Sherif et al. been so influential?
2. How important is group conflict as a cause of prejudice?

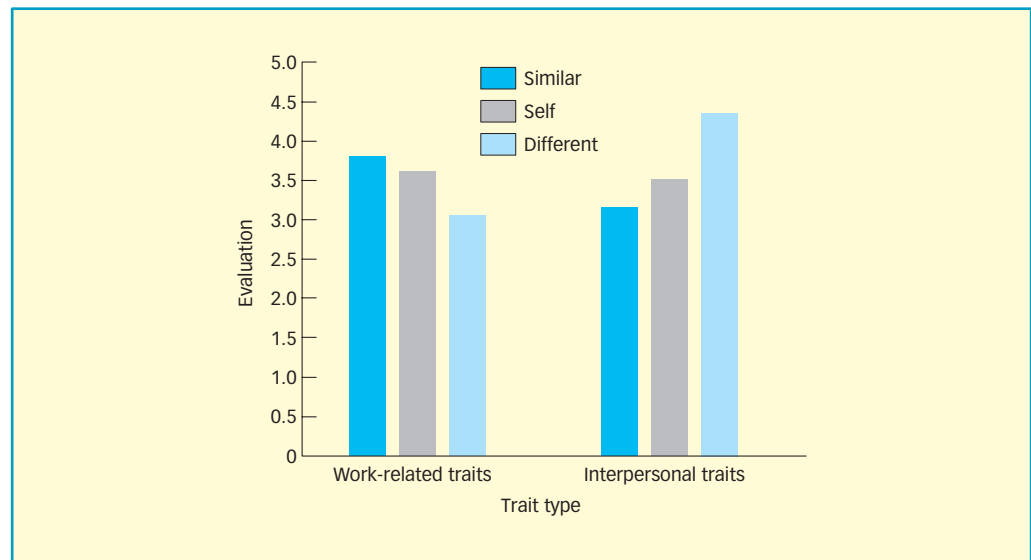
KEY STUDY EVALUATION

Sherif et al.'s study has been regarded as very important because it showed ordinary boys acting in different ways toward each other depending on the situation. Competition resulted in dislike and hostility; a common goal led to friendship and good feelings. It might be interesting to speculate about whether the results would have been different if all the participants had been girls. It has been argued that while they are growing up girls are rewarded for cooperation, whereas boys are rewarded for competitiveness. It could also be argued that the participants were not a representative group, in that they were not randomly selected.

Ember (1981) studied 26 small societies. As predicted by realistic group conflict theory, intergroup violence was much more frequent when societies had to compete for resources because of population pressures or severe food shortages.

Zárate, García, Garza, and Hitlan (2004) studied prejudice against Mexican immigrants at the University of Texas at El Paso, which is on the border between the United States and Mexico. They predicted from realistic group conflict theory that prejudice against Mexican immigrants would be increased if participants focused on similarities between the two groups in work-related traits. The findings supported this prediction (see figure on the following page), presumably because this manipulation threatened participants' sense of job security. There was also evidence for the importance of cultural threat. When differences between Mexican immigrants and native Americans on interpersonal traits were emphasized, this led to increased prejudice.

Mean evaluation of an outgroup (Mexican immigrants) as a function of participants having previously rated how similar or different the outgroup was to their ingroup with respect to work-related traits or interpersonal traits (the self condition is a control condition). High evaluation scores indicate more prejudice than low scores. From Zarate et al. (2004). Copyright © 2004 Elsevier. Reproduced with permission.



Evaluation

- + Competition between two groups for the same goal can lead to prejudice.
- + Realistic conflict theory helps to explain the large increases in prejudice found when countries are at war with each other.
- According to the theory, conflicts arise when group interests are threatened. However, group interests are defined very vaguely: “A real or imagined threat to the safety of the group, an economic interest, a political advantage, a military consideration, prestige, or a number of others” (Sherif, 1966, p. 15).
- Realistic conflict is not always sufficient to produce prejudice (e.g., Tyerman & Spencer, 1983). It is also not necessary, because millions of people have prejudiced attitudes toward people in other cultures with whom they are not in competition. Additional factors such as cultural threat also need to be considered (e.g., Zárate et al., 2004).

SOCIAL IDENTITY THEORY

According to social identity theory (discussed earlier), individuals seek to distinguish their ingroup as clearly as possible from outgroups. They typically regard their ingroup favorably in comparison with outgroups to boost their self-esteem. As a result, members of outgroups may be exposed to prejudice and discrimination.

Doosje, Branscombe, Spears, and Manstead (1998) considered the attitudes of Dutch people toward Indonesia (previously a Dutch colony). Some participants had a high level of national identification with being Dutch, whereas others had only a low level. As predicted by social identity theory, those having the greatest ingroup bias (i.e., high level of national identification) felt less guilty about Dutch treatment of the Indonesians. They were also less willing to compensate the Indonesians than were those having little ingroup bias (i.e., low level of national identification).

Verkuyten, Drabbles, and van den Nieuwenhuijzen (1999) tested predictions of social identity theory in Dutch participants aged between 16 and 18. These participants indicated how strongly they identified themselves with the majority ingroup in Holland. Their level of prejudice was assessed by asking them to indicate their emotional reactions to descriptions of situations involving ethnic minorities (e.g., “More and more Islamic

people who have different views and habits have come to live in the Netherlands. Because of this there is an increasing danger that the Dutch norms and values are threatened”). Individuals identifying themselves most strongly as Dutch revealed the greatest negative emotions toward ethnic minorities. Thus, group identification was related to prejudice as predicted by social identity theory.

In sum, social identity theory helps to account for individual differences in prejudice. However, it provides a limited view, because it ignores important factors such as intergroup competition and conflict. Another limitation is that the theory indicates why prejudice and social conflict exist, but doesn't make it clear in detail *how* they develop.

REDUCING PREJUDICE AND DISCRIMINATION

What can be done to reduce (and ideally eliminate) prejudice and discrimination? Psychologists have provided several answers to that question, and some of the main ones are discussed here.

INTERGROUP CONTACT HYPOTHESIS

Allport (1954) argued in his intergroup contact hypothesis that the most effective way of reducing prejudice is by intergroup contact. Four conditions need to be met if such contact is to prove successful:

1. The two groups have equal status within the situation in which the contact takes place.
2. The two groups work toward common goals. Efforts to achieve these common goals are based on intergroup cooperation.
3. Intergroup contact must occur often (and long) enough to permit the development of meaningful relationships between the members of the different groups.
4. There should be social and institutional support for intergroup acceptance.

Wright, McLaughlin-Volpe, and Ropp (1997) developed some of Allport's ideas. They found evidence for the **extended contact effect**: prejudice is reduced when a member of your ingroup is very friendly with a member of the other group.

Findings

Aronson and Osherow (1980) obtained support for the intergroup contact hypothesis. Schools in Austin, Texas had recently been desegregated, and there were concerns about racial conflict. One class of black and white children was divided into small groups for a learning task (e.g., the life of Abraham Lincoln). Within each group, every child learned a different part of the information. Each group member then taught what he/she had learned to the others. After that, the children received a mark based on their overall topic knowledge. This approach was called the **jigsaw classroom**, because all the children contributed just as all the pieces of jigsaw are needed to complete it.

The findings from the jigsaw classroom were promising. The children showed higher self-esteem, better



Having children work in groups helps reduce racial barriers.

Key Terms

Extended contact effect:

knowledge that a member of your ingroup has formed a close relationship with a member of another group can reduce prejudice against that other group.

Jigsaw classroom:

an approach to teaching designed to reduce prejudice in which the teacher ensures that all children in the class contribute to the achievement.

Case Study: *New Era Schools Trust*

The New Era Schools Trust (or NEST) runs three boarding schools in South Africa, in Durban, Johannesburg, and Cape Town. The unique aim of all the NEST schools is not only to produce well-educated and personable young people, but also to eliminate any trace of racial prejudice in their students. To achieve this, all races are mixed together from the very first day at school, living and studying alongside each other in a way that is rare even in post-apartheid South Africa. The teachers are similarly multiracial, and there is an equal mix of boys and girls.

Not only are the different races regarded as equal in NEST schools, their cultures are also given equal value. Schools in South Africa have generally taken the view that African culture is irrelevant, and have taught exclusively from a white perspective. At NEST schools the pupils study Xhosa poets as well as Keats, and the lives of Zulu warriors as well as Napoleon. This sense of total equality permeates everything—there are no prefects or top-down discipline, no uniforms or corporal punishment, and everyone takes a hand in doing the chores.

NEST has found that more black parents than white parents wish their children to attend a NEST school. White children tend to have better access to well-equipped schools where they are not required to help clean the dormitories, whereas many black parents are keen for their children to leave the deprivation of the townships to receive their education. This imbalance is lessening, however, as white parents realize what good academic success the NEST schools are achieving. In 1992 their pass rate was 100%, when private white schools and white church schools averaged 90%.

(Based on an article by Prue Leith, *The Times*, May 1993.)

school performance, more liking for their classmates, and reduced prejudice. However, these effects were small, probably because the jigsaw classroom was only used for 12 hours spread over 6 weeks.

Rosenfield, Stephan, and Lucker (1981) identified problems with the jigsaw classroom. Minority group members low in competence were blamed for slowing down the learning of the more competent students. This confirmed existing prejudiced attitudes rather than reducing them.

A thorough test of the intergroup contact hypothesis was carried out at Wexler Middle School in Waterford in the United States (see Brown, 1986). The number of black and white students was similar so that it wouldn't be regarded as a black or a white school. Much was done to make all the students feel equal, with very little streaming on the basis of ability. Cooperation was encouraged by having the students work together to buy special equipment for everyone.

The results over the first 3 years were moderately encouraging. There were many black–white friendships, but they rarely extended to visiting each other's homes. There was a steady reduction in discrimination, with the behavior of the black and white students toward each other being friendly. However, some stereotyped beliefs remained. Black and white students agreed that black students are tougher and more assertive than white students, and that white students are cleverer and work harder.

Tropp and Pettigrew (2005) pointed out that most studies on intergroup contact have focused only on the effects of such contact on the majority group. Tripp and Pettigrew found by carrying out meta-analyses that the beneficial effects of intergroup contact are greater for majority than for minority groups. The correlation between amount of intergroup contact and prejudice was $-.23$ for majority groups compared to only $-.18$ for minority groups. Tropp and Pettigrew found that the beneficial effects of intergroup contact on majority groups were especially large when the conditions under which that contact occurred were those regarded as optimal by Allport (1954). However, this was *not* the case for minority groups. Members of minority groups may be so aware of their lower status that it reduces any positive effects of intergroup contact.

Evaluation

- + The intergroup contact hypothesis identified several factors influencing whether intergroup contact will reduce prejudice.
- + Much evidence provides general support for the hypothesis in majority groups. However, while it is fairly easy to reduce prejudice to individual members of another group, it is much more difficult to ensure this *generalizes* to all members of the other group.
- It is difficult to interpret the finding that high levels of intergroup contact are associated with low levels of prejudice within the majority group. One possibility is that prejudiced individuals do their best to avoid intergroup contact.
- The intergroup contact hypothesis has little to say about *how* or *why* contact reduces prejudice, and it doesn't indicate how positive contact with individual members of an outgroup might generalize to include other members of that outgroup.
- Intergroup contact typically has little effect on prejudice within minority groups (Tropp & Pettigrew, 2005).

SALIENT CATEGORIZATION AND RECATEGORIZATION

Suppose someone prejudiced against an outgroup has positive social interactions with a member of that outgroup. Such interactions may well lead to liking for that individual. However, this often doesn't lead to reduced prejudice toward the outgroup as a whole. We will now consider two theoretical approaches designed to solve this problem (and other issues relating to prejudice).

Salient categorization

Why do positive experiences with one member of an outgroup often fail to generalize to other members of that outgroup? According to Hewstone and Brown (1986), the group member is generally regarded as “an exception to the rule.” What is needed is to make the outgroup member's group membership salient or obvious so that he/she is perceived as typical or representative of his/her group. Thus, salient categorization may provide the key. Salient categorization is an important ingredient in the multicultural approach, according to which group differences should be acknowledged and celebrated.

According to social identity theorists (see earlier in the chapter) emphasizing group differences has to be handled sensitively if it is to reduce prejudice. For example, individuals typically distinguish between an ingroup and an outgroup, and show ingroup bias or favoritism that can easily turn into prejudice. There is a danger that salient categorization may simply serve to *increase* prejudice against other groups.

Van Oudenhoven, Groenewoud, and Hewstone (1996) tested the notion that it is important for the group membership of an outgroup member to be salient or obvious if prejudice is to be reduced. Dutch participants spent 2 hours interacting with a Turkish confederate in one of three conditions:

1. The experimenter never mentioned the confederate's ethnicity (low salience).
2. The experimenter only mentioned the confederate's ethnicity at the end of the 2-hour period (moderate salience).
3. The confederate's ethnicity was emphasized throughout the session (high salience).

Attitudes toward Turks in general were more favorable in the moderate and high salience conditions than in the low salience condition.

Brown, Vivian, and Hewstone (1999) carried out a study in which British participants worked with a German confederate to obtain a substantial reward. The German

confederate either seemed to correspond fairly closely to the typical German stereotype or was clearly atypical. In addition, participants were given false information about how similar or dissimilar German people are with respect to several characteristics. It was assumed that it would be easiest to generalize from the German confederate to all Germans when the confederate was regarded as typical and when the Germans were thought to be similar to each other. As predicted, participants in that condition had the most favorable attitudes toward the Germans.

Voci and Hewstone (2003) studied prejudice toward immigrants shown by Italian hospital workers. They found that the effects of intergroup contact on prejudice depended very much on group salience. As predicted, reduced prejudice occurred most frequently when hospital workers had positive contact with immigrants, and the group to which immigrants belonged was salient. Positive contact plus group salience reduced anxiety about immigrants, which then led to reduced prejudice.

Evaluation

- + Salient categorization has been shown experimentally to reduce prejudice to entire outgroups rather than simply to specific individuals.
- + Much nonexperimental research in education has found that multiculturalism (which resembles salient categorization) improves intergroup relations more than the alternative color-blind or recategorization approach (see Richeson & Nussbaum, 2004, and next section).
- When people already possess a clear stereotype of another group, it may be very difficult to persuade them that a given individual is actually a typical member of that group.
- “If the cooperative interaction goes wrong . . . then structuring the interaction at the intergroup level could make matters worse . . . there is a risk of reinforcing negative stereotypes of the outgroup precisely because those people are seen as typical of it” (Brown, 2000, p. 353).

Recategorization

As we saw earlier, there is evidence of ingroup bias in which individuals regard their own group as superior to other groups. This phenomenon suggests a way of dealing with prejudice. Prejudice could be eliminated by a process of **recategorization** in which the ingroup and outgroup are recategorized to form a single ingroup. Precisely this was proposed by Gaertner et al. (1993). Recategorization is of central importance with the

Television and the reduction of prejudice

The Children’s Television Workshop (CTW) does more than make the *Sesame Street* series. In a research project with Israel Educational Television and Al-Quds University’s Institute for Modern Media, in east Jerusalem they made a similar set of television programs for preschoolers aimed at Israeli and Palestinian children (*Rechov Sumsum/Shara’a Simsim* Research Symposium, 1999). Prejudiced attitudes toward the other culture had been found in children as young as 4 years old. The agenda of the programs was to emphasize similarities between children in the two cultures. The series *Rechov Sumsum/Shara’a Simsim* (Hebrew and Arabic translation, respectively) was broadcast to viewers in Israel and the West Bank in 1998. A follow-up study in 1999 showed a small but real reduction in the prejudiced attitudes of child viewers.

Key Term

Recategorization: the process of producing a new categorization in which the ingroup and the outgroup are combined into a single ingroup.

color-blind perspective, according to which category memberships should be reduced or eliminated.

Gaertner et al. (1994) studied recategorization in a multi-ethnic high school in the United States. They carried out a survey, part of which consisted of items focusing on the notion that there was a single ingroup within the school. Those students who thought the school consisted of one large ingroup had the most positive attitudes toward other ethnic groups in the school.

Dovidio, Gaertner, and Validzic (1998) had two groups working together on a task. These groups were equal or unequal in status, and their area of expertise was either the same or different. There was most recategorization of the two groups as a single ingroup when the two groups were equal in status and their expertise differed. As predicted, this was the *only* condition in which outgroup prejudice was eliminated. Dovidio et al. speculated that intergroup contact reduces prejudice mainly when it serves to produce recategorization.

Individuals often exhibit much more prejudice than was shown in the study by Dovidio et al. (1998). A test of the importance of recategorization under more realistic conditions was conducted by Dovidio et al. (2004). White participants watched a videotape showing examples of racial discrimination, followed by assessment of their level of prejudice. Those who had previously been exposed to a recategorization manipulation (told that a terrorist threat was directed at black and white Americans) showed reduced prejudice compared to another group told that a terrorist threat was directed only at white Americans.

Recategorization can involve some loss of social identity as an individual's ingroup is combined into a larger group, and this may have negative consequences. Evidence that this is a danger was reported by Crisp, Stone, and Hall (2006), who divided their participants into those with low and those with high ingroup identification. Participants who identified with their ingroup showed evidence of *increased* prejudice following recategorization.

Evaluation

- + Recategorization has been shown to produce beneficial effects including prejudice reduction whether the initial prejudice was high (e.g., Dovidio et al., 2004) or low (e.g., Dovidio et al., 1998).
- It is difficult to change perceptions of an outgroup in the face of pre-existing strong and habitual prejudice, as is often the case in the real world.
- Individuals may perceive that any potential gains from recategorization are outweighed by the loss of their social identity stemming from their current ingroup membership (e.g., Crisp et al., 2006).

OVERALL SUMMARY

Intergroup contact often reduces prejudice within majority groups under the appropriate conditions (e.g., the two groups are of equal status within the situation; meaningful relationships are developed). Two mechanisms that can enhance the beneficial effects of intergroup contact are salient categorization and recategorization, both of which stem from the social identity approach. However, there are limitations and dangers associated with use of both mechanisms, and salient categorization can increase ingroup bias and thus prejudice. It is important for future research to discover *why* intergroup contact generally fails to reduce prejudice within minority groups, and to use that knowledge to devise ways of remedying the situation.

Chapter Summary

Social identity

- According to social identity theory, social identities make us feel good about ourselves and enhance our self-esteem.
- Ingroup bias or favoritism has been found with the allocation of both positive and negative outcomes.
- Ingroup bias increases the self-esteem of members of the ingroup. However, there is no tendency for individuals who initially have low self-esteem to exhibit more ingroup bias than those with high self-esteem.
- Social identity theory focuses on the cognitive processes underlying group identification and de-emphasizes emotional and motivational factors.

Stereotypes

- Contrary to what is often believed to be the case, there is a grain of truth in most stereotypes.
- Problems with questionnaire assessment of stereotypes (e.g., social desirability bias) can be overcome by assessing implicit stereotypes or attitudes.
- According to the cognitive approach, stereotypes reduce the amount of processing needed when we meet or think about other people. However, stereotypes are used much more flexibly than is assumed within this approach.
- According to the social approach, the stereotypical information we focus on depends on the prevailing social context. However, most stereotypes have an unchanging central core of meaning.
- It is difficult to change stereotypes because individuals violating the stereotype of their group are regarded as unrepresentative and are assigned to a separate subtype.

Prejudice and discrimination

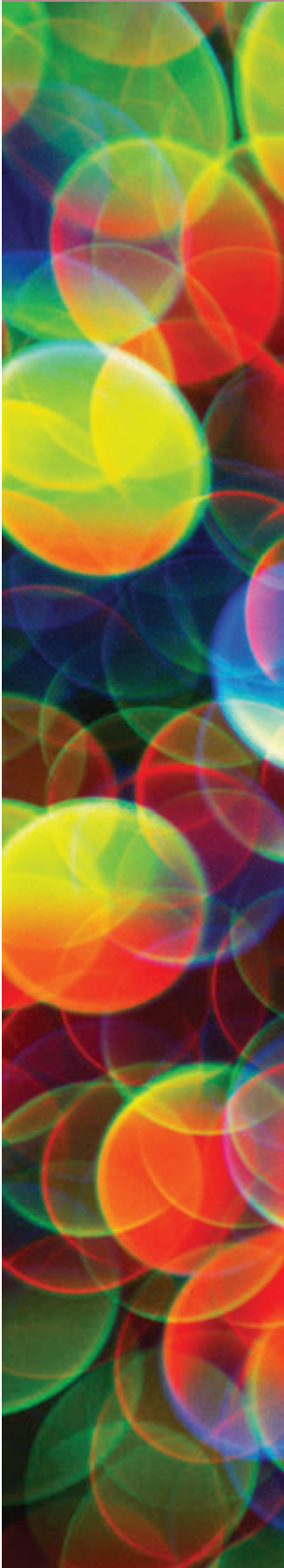
- There is only a modest association between prejudice and discrimination, because there are greater social pressures to avoid discrimination.
- Millions of people in Western societies exhibit aversive or ambivalent racism, which involves a conflict between egalitarian values and negative emotions.
- According to role congruity theory, women occupy few top positions for two reasons. First, they are regarded as not possessing the qualities needed for leadership. Second, decisive leadership is regarded less favorably when shown by a woman than by a man.
- According to Adorno et al. (1950), individuals with an authoritarian personality are prejudiced because they displace hostility toward their parents onto minority groups. In fact, however, prejudice depends more on cultural norms than on personality.
- According to realistic group conflict theory, prejudice results from intergroup conflict. There is support for this theory, but intergroup conflict is neither necessary nor sufficient to produce prejudice.
- According to social identity theory, individuals showing the greatest ingroup bias should be most prejudiced against outgroups. There is support for this prediction, but the theory ignores other determinants of prejudice (e.g., intergroup conflict).

Reducing prejudice and discrimination

- According to Allport's intergroup contact hypothesis, intergroup contact between two groups of equal status working toward common goals produces reduced prejudice.
- There is reasonable support for the intergroup contact hypothesis. However, support is stronger for majority groups than for minority ones, and it may mainly be people who are already unprejudiced who seek intergroup contact.
- Producing a reduction in prejudice is often easier when an outgroup member's group membership is made salient or obvious.
- Prejudice can be reduced by recategorization, in which the ingroup and the outgroup are recategorized to form a single ingroup. However, there can be problems if group members have a sense of loss of their social identity.

Further Reading

- Brewer, M.B. (2003). *Intergroup relations* (2nd ed.). Philadelphia: Open University Press. This is a wide-ranging book that includes much coverage of prejudice and prejudice reduction.
- Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, 53, 575–604. This paper contains a good account of various theoretical accounts of intergroup bias and conflict.
- Hogg, M.A., & Vaughan, G.M. (2005). *Social psychology* (4th. ed.). Harlow, UK: Prentice Hall. Chapters 10 and 11 of this excellent textbook provide clear discussions of intergroup phenomena including prejudice.
- Wright, S.C., & Taylor, D.M. (2003). The social psychology of cultural diversity: Social stereotyping, prejudice, and discrimination. In M.A. Hogg & J. Cooper (Eds.), *The Sage handbook of social psychology*. London: Sage. All of the main topics dealt with in this chapter are discussed thoroughly by Wright and Taylor in this book.



- **Approaches to abnormality** pp.511–521

What is abnormality?

Classification systems

Factors causing mental disorders

- **Therapeutic approaches** pp.523–537

Drug therapy

Psychodynamic therapy

Behavior therapy

Cognitive and cognitive-behavior therapy

Effectiveness of therapy

How effective are different forms of therapy?

INTRODUCTION TO Abnormal Psychology

W

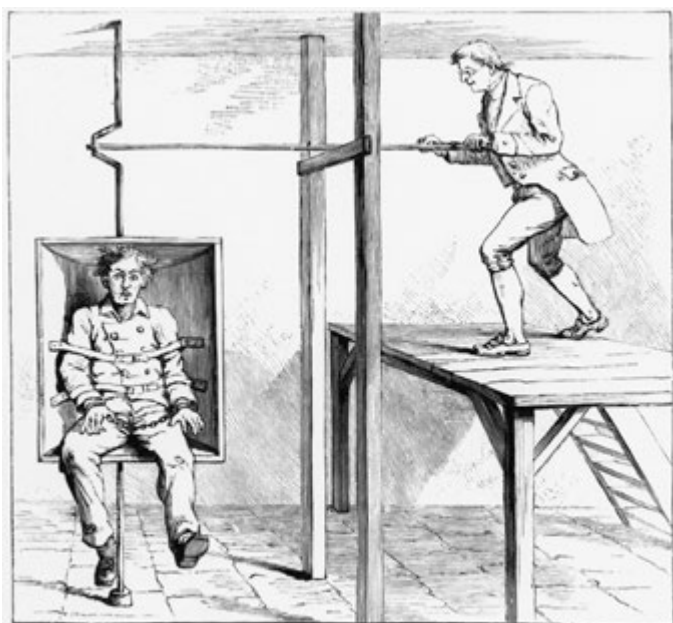
e live in an era in which huge numbers of people struggle to cope with various serious psychological problems. In the United States, as Comer (2001, p. 7) pointed out:

Up to 10 of every hundred adults have a significant anxiety disorder, 10 suffer from profound depression, 5 display a personality disorder . . . 1 has schizophrenia . . . and 11 abuse alcohol or other drugs. Add to these figures as many as 600,000 suicide attempts, 500,000 rapes, and 3 million cases of child abuse each year, and it becomes apparent abnormal psychological functioning is a pervasive problem.

Twenge (2000) assessed trait anxiety (tendency to experience much anxiety) among American children and college students. The mean score for both groups was much higher than used to be the case. This led Twenge (2000, p. 1007) to conclude that, “The average American child in the 1980s reported more anxiety than child psychiatric patients in the 1950s.” This suggests that we live in a world that is becoming ever more stressed. I am not so sure. It is probably truer to say that the nature of life’s stressors has changed somewhat over the past century or so. As little as 100 years ago, for example, life expectancy in the Western world was 30 years less than now. As a result, young and middle-aged people in those days were probably much more worried and stressed about their physical health than is the case today. We may worry about how to find the money for a good holiday somewhere in the sun. However, 100 years ago most people couldn’t afford to have any holidays at all, and it was often a struggle to put food on the table.

The focus of this chapter is on abnormal psychology. The achievements of psychologists in this area can be seen very clearly by considering history. For many centuries, the treatment applied to those suffering from mental disorders was positively barbaric. It was believed that mental disorders were caused by demons or other supernatural forces. Popular “cures” for mental illness involved making things as unpleasant as possible for the demons. The techniques used included immersing the patient in boiling hot water, flogging, starvation, and torture. It was thought these “cures” would persuade the demons to leave the patient’s body and so remove his/her disorder.

As is well-known, Sigmund Freud (at the end of the nineteenth century) was the first psychologist to argue strongly that psychological approaches to treatment were needed (see Chapter 2). He was also the first psychologist to provide a detailed and systematic therapeutic approach by developing psychoanalysis (discussed in detail later). As a result of these contributions (and numerous others in many areas of psychology), he is deservedly the most famous psychologist of all time.



THE DOCTOR THINKS THAT "NO WELL-REGULATED INSTITUTION SHOULD BE UNPROVIDED WITH THE CIRCULATING SWING." 1818.

An illustration of an early nineteenth century swing used to treat depression. The "circulating swing" was supposed to bring the depressive "back to sound reasoning."

If we go back before the time of Sigmund Freud, it is worth mentioning the Austrian mystic and physician Franz Mesmer (1734–1815). He treated patients suffering from various complaints by sitting them around a tub containing magnetized iron filings with protruding iron rods. It was claimed that cures were produced by the "animal magnetism" generated by this bizarre arrangement. Subsequently, however, it became clear that the sleep-like or hypnotic state involved in the exercise (rather than any animal magnetism!) was responsible for the cures that were achieved. The importance of Mesmer's work is that he showed (although he didn't realize it at the time!) that mental disorders could be cured by psychological techniques.

Chapters 21 and 22 are devoted to the key issues relating to mental disorders. These key issues can be expressed in the form of five questions. First, how can we best describe and categorize mental disorders? Second, what are the major psychological approaches that have been applied to mental disorders in recent decades? Third, what are the factors responsible for the development of the various mental disorders? Fourth, what are the main forms of psychological therapy that have been used to treat individuals with mental disorders? Fifth, how can we assess the effectiveness of different forms of psychological therapy in producing beneficial changes in individuals suffering from mental disorders?

We can illustrate the above issues by considering a hypothetical individual, Matthew. Matthew is extremely shy, and dislikes most social occasions. Indeed, he is so frightened of possible humiliation when involved in a social situation that he will often make what sound to other people like feeble excuses to avoid them. He goes to see a clinical psychologist, because he is so concerned about his condition.

The successful treatment of Matthew's conditions involves several steps. First, we need to decide the precise nature of the problem. Thus, diagnosing his condition is important. Second, it may be useful to consider Matthew's condition within the context of the main models or theoretical approaches to abnormal psychology. Third, we need to combine our diagnosis of his condition with an appropriate model to appreciate the factors responsible for his problem. Fourth, we need to treat Matthew to eliminate his symptoms and prevent them from recurring. Fifth, we need to use various measures (e.g., assessing his mood state, ability to function in everyday life, and his general behavior) to reassure ourselves that the treatment has been effective.

Chapter 21 deals with the various psychological problems and disorders that individuals can experience. This area is also known as **psychopathology**, which is "the field concerned with the nature and development of mental disorders" (Davison & Neale, 1998, p. G-19). More than 200 mental disorders have been identified, and most (but not all) of these disorders are found in virtually every country in the world. More specifically, Chapter 21 provides answers to the first three questions posed above. In other words, it is concerned with issues such as the diagnosis of mental disorders, identifying the main theoretical approaches to abnormality, and developing an understanding of the factors responsible for the main mental disorders.

The emphasis in Chapter 22 is very different from the one adopted in Chapter 21: it is concerned with *practical* issues relating to the forms of therapy provided by clinical psychologists, psychiatrists, and others. It is over 100 years since Freud first proposed a systematic form of treatment for mental disorders, since when numerous other forms of treatment have been devised. Chapter 22 answers the fourth and fifth questions posed earlier. That is to say, there is a detailed discussion of the major forms of therapy associated with each of the main approaches to abnormality. After this has been

Key Term

Psychopathology: the study of the nature and development of mental disorders; an abnormal pattern of functioning.

accomplished there is an attempt to evaluate the effectiveness of these forms of treatment. As we will see, it is surprisingly difficult to come to definitive conclusions concerning the relative effectiveness of different forms of therapy. However, the good news is that all of the main types of treatment have been shown beyond any doubt to be at least moderately effective in the treatment of a wide range of mental disorders. There is also some evidence that certain forms of treatment are especially successful and effective when applied to specific mental disorders.

chapter 21

Contents

What is abnormality?	511
Classification systems	512
Factors causing mental disorders	514

Approaches to abnormality

21

WHAT IS ABNORMALITY?

It is more difficult than you might imagine to decide whether a given individual is abnormal or suffers from a mental disorder. Why is that? The central problem is that concepts such as “abnormality” and “mental disorder” are vague. As Lilienfeld and Marino (1999) pointed out, the concept of “mental disorder” has fuzzy boundaries, so that “there is no . . . set of criteria in nature that can be used to definitively distinguish all cases of disorder from all cases of nondisorder” (p. 400).

One reason we cannot be precise about these issues is because they depend in part on the prevailing social norms and values. For example, consider changing views about homosexuality as reflected in a major classificatory system known as the Diagnostic and Statistical Manual of Mental Disorders (DSM). In DSM-II, published in 1968, homosexuality was classified as a mental disorder involving sexual deviation. By 1987, with the publication of DSM-III-R (the revised third edition), only homosexuals having “persistent and marked distress about their sexual orientation” were regarded as having a disorder.

FOUR DS

In spite of the vagueness of the concept of “abnormality” or “mental disorder,” we can identify features often (but not invariably) associated with it. Comer (2001) argued there are four central features (known as the four Ds):

- *Deviance*: This involves thinking and behaving in ways not regarded as acceptable within a given society. According to Comer (2001, p. 3), “Behavior, thoughts, and emotions are deemed abnormal when they violate a society’s ideas about proper functioning. Each society establishes norms—explicit and implicit rules for proper conduct . . . Behavior, thoughts, and emotions that violate norms of psychological functioning are called abnormal.”
- *Distress*: It is not sufficient for behavior to be deviant for it to be regarded as abnormal. For example, Comer (2001) pointed out that there are people in Michigan called the Ice Breakers, who go swimming in extremely cold lakes every weekend between November and February. This behavior violates society’s norms. However, it wouldn’t generally be regarded as abnormal, in part because the Ice Breakers experience no distress. Thus, deviant thoughts and behavior need to cause distress to the individual and/or others to be considered abnormal.
- *Dysfunction*: According to Comer (2001, p. 4), “Abnormal behavior tends to be dysfunctional; that is, it interferes with daily functioning. It so upsets, distracts, or confuses people that they cannot care for themselves properly, participate in ordinary social interactions, or work productively.” Most people exhibit dysfunctional behavior in this sense when bereaved, but the duration of such behavior is less than in most cases of abnormality.



- **Danger:** Individuals whose behavior poses a threat or danger to themselves or to others are generally regarded as abnormal. However, most individuals having a mental disorder do *not* pose a danger to anyone, and some dangerous individuals (e.g., armed criminals) don't suffer from a mental disorder.

Comer (2001, p. 5) concluded that the four-D approach is an imprecise way of defining or identifying abnormality: “While we may agree to define psychological abnormalities as patterns of functioning that are deviant, distressful, dysfunctional, and sometimes dangerous . . . these criteria are often vague and subjective. When is a pattern of behavior deviant, distressful, dysfunctional, and dangerous enough to be considered abnormal? The question may be impossible to answer.”

CLASSIFICATION SYSTEMS

Several classification systems are used to diagnose mental disorders. For example, we have already mentioned the Diagnostic and Statistical Manual of Mental Disorders (DSM), which contains over 200 mental disorders. Another prominent system is the International Classification of Diseases (ICD). This is produced by the World Health Organization, and is much used throughout Europe and other parts of the world. The two classificatory systems are broadly similar, and so we will focus on only one of them (DSM).

The current version of DSM is DSM-IV, published in 1994 (a slightly revised version known as DSM-IV-TR (text revision) was published in 2000). It consists of five axes. The first three are always used, whereas the last two are optional:

- Axis 1: *Clinical disorders.* This axis permits the patient's disorder to be diagnosed on the basis of symptom patterns.
- Axis 2: *Personality disorders and mental retardation.* This axis identifies long-term patterns of impaired functioning stemming from personality disorders or mental retardation.
- Axis 3: *General medical conditions.* This axis concerns any physical illness that might influence the patient's emotional state or ability to function effectively.
- Axis 4: *Psychosocial and environmental problems.* This axis is concerned with any significant stressful events occurring within 12 months of the onset of the mental disorder.
- Axis 5: *Global assessment of functioning.* This axis provides an overall measure of the patient's functioning at work and at leisure on a 100-point scale.

It has been argued that DSM suffers from two major problems. First, it is an American-based system, and it fails to take proper account of cultural factors. Second, a persistent criticism of DSM is that it is flawed by sex bias. More specifically, it is claimed that male-biased assumptions about what should be regarded as abnormal have influenced the diagnostic categories used in DSM. We consider these issues in turn.

CULTURAL DIFFERENCES

DSM-IV contains a modest attempt to take account of cultural factors. It refers to **culture-bound syndromes**, which are “locality-specific patterns of aberrant [deviant] behavior and troubling experience that may or may not be linked to a particular DSM-IV diagnostic category.” Here are three examples of the culture-bound syndromes identified by DSM-IV in an appendix (other examples are shown in the box opposite):

- **Ghost sickness.** The main symptom is an excessive focus on death and on those who have died (common in Native American tribes).
- **Koro.** This disorder involves extreme anxiety that the penis or nipples will recede into the body, and possibly cause death (south and east Asia).
- **Amok.** This disorder involves brooding followed by a violent outburst; it is found mainly in men (originally identified in Malaysia).

Key Term

Culture-bound syndromes: patterns of disordered behavior typically found in only certain cultures.

Some examples of culture-bound syndromes

Country	Syndrome	Key features
Caribbean	Blackening out	Sudden fainting + hysterical blindness
China	Pa-feng	Fear of wind
Greece	Nevra	Emotional distress, stomach complaints, dizziness, and so on
Japan	Taijin kyofusho	Extreme fear that one's body or body parts are offensive to other people
Latin America	Mal de ojo	The "evil eye," responsible for behavioral problems and poor health
South Africa	Amafufunyana	Violent behavior caused by spirit possession
West Africa	Brain fag	Problems in concentrating and thinking produced by excessive study (!)

Case Study: A Culture-Bound Syndrome?—*Hikikomori*

A condition attracting considerable concern in Japan over recent years is *Hikikomori*. There has been no official calculation of the number of cases, but some specialists have estimated that it may be affecting up to a million people. The condition affects mostly middle-class males in their late teens or early twenties who are otherwise healthy.

Sufferers withdraw completely from society, typically by locking themselves in their rooms, sometimes for up to 20 years. There have been some high-profile cases reported in the Japanese media where young men have left their

homes and committed violent crimes, including murder. However, most of the sufferers are not violent and tend more towards depression and lethargy. Other symptoms can include insomnia, regressive behavior, paranoia and aspects of agoraphobia and obsessive-compulsive disorder.

The Japanese government is of the opinion that *Hikikomori* is a social disorder rather than a mental disorder, and that it reflects the current economic downturn in the country. (In a similar way, *Karoshi*—death from overwork—was a symptom of Japan's huge economic success in the 1990s.)

Kleinman and Cohen (1997, p. 76) dismissed the appendix of DSM-IV as "little more than a sop thrown to cultural psychiatrists and psychiatric anthropologists." This is fair comment when we consider the range of culture-bound syndromes around the world.

SEX BIAS

Is there sex bias in DSM? Many argue that there is focus on various personality disorders. For example, it is claimed that histrionic personality disorder (characterized by excessive emotionality) represents distortions of stereotypical feminine traits. In contrast, antisocial personality disorder (characterized by aggression and hostility) represents distortions of stereotypical masculine traits. Evidence suggesting the existence of sex bias was reported by Ford and Widiger (1989). Therapists were presented with written case studies of patients with each of those personality disorders. Histrionic personality disorder was correctly diagnosed much more often when the patient was described as female than when described as male (80% vs. just over 30%). Antisocial personality disorder was correctly diagnosed over 40% of the time when the patient was male but under 20% of the time when the patient was female.

As Funtowicz and Widiger (1999) pointed out, the findings of Ford and Widiger (1989) may reflect sex bias in the DSM criteria or they may be a result of biases in the therapists making the diagnoses. Funtowicz and Widiger argued that sex bias would be shown if the criteria for diagnosing personality disorders occurring more often in females are lower than those for personality disorders occurring more often in males. They asked clinical psychologists to indicate the extent to which the criteria for DSM-IV personality disorders involve social dysfunction, occupational dysfunction, or subjective distress. The diagnostic criteria were as stringent for "female" personality disorders as for "male" ones, thus providing no evidence of sex bias.

Evaluation

- + DSM is a reliable and effective classificatory system.
- + DSM seems to be free of sex bias, although such bias may occur when therapists use it in practice.
- DSM-IV takes insufficient account of culture-bound syndromes.
- DSM-IV makes use of categories, i.e., patients have (or don't have) a given mental disorder. With a dimensional system, in contrast, it is assumed that individuals vary in the extent to which they possess various symptoms. Since individuals have varying degrees of symptoms such as anxiety and depression, a dimensional approach is less artificial than a categorical one.
- According to Pilgrim (2000), *all* classification systems (including DSM-IV) are based too much on symptoms and ignore the personal and social context in which those symptoms occur.

FACTORS CAUSING MENTAL DISORDERS

It is important to establish the causes of mental disorders, in part because the knowledge gained could be used to reduce the incidence of such disorders in future. In this section, we consider etiology, which is the range of factors causally involved in the development of mental disorders.

It is a complex matter to find out *why* some people suffer from any given mental disorder. However, we can start by distinguishing between one-dimensional and multi-dimensional causal models (Durand & Barlow, 2006). According to one-dimensional models, the origins of a mental disorder can be traced to a *single* underlying cause. For example, severe depression might be caused by a major loss (e.g., death of a loved one) or schizophrenia by genetic factors. In practice, one-dimensional models have consistently failed to account for the evidence. Accordingly, they have been replaced by multidimensional models assuming that any given mental disorder is typically caused by *several* factors in interaction.

Probably the most popular multidimensional approach is the **diathesis–stress model**. In this model, the occurrence of mental disorders depends on two types of factors:

1. **Diathesis.** This is a vulnerability or predisposition to disorder within the individual. Diathesis used to be thought of simply as a genetic vulnerability, but has broadened to include any personal vulnerabilities.
2. **Stress.** This is some severe or disturbing environmental event (e.g., divorce; death of a spouse).

The main types of factors identified as contributing to the development of mental disorders are as follows:

- **Genetic factors.** Twin studies, family studies, and adoption studies may indicate that some people are genetically more vulnerable than others to developing a disorder.
- **Brain chemicals.** Individuals with unusually low or high levels of certain brain chemicals may be vulnerable to various mental disorders.
- **Cultural factors.** Cultural values and expectations may be important in causing some disorders. For example, most Western cultures emphasize the desirability of thinness in women, and this may help to trigger eating disorders.

Key Term

Diathesis–stress model: the notion that mental disorders are caused jointly by a diathesis or vulnerability within the individual *and* a distressing event.



Death of a loved one rates as one of the most stressful life events we experience, and may have long-term psychological repercussions, such as depression.

- **Social factors.** Individuals experiencing severe life events (e.g., divorce; unemployment) may be at risk for various psychological disorders, as may those lacking social support or belonging to poorly functioning families.
- **Psychological factors.** Individuals having certain kinds of beliefs, or who have learned particular inappropriate ways of behaving, may be vulnerable to mental disorders. For example, someone who exaggerates the threat of most situations may develop an anxiety disorder.

The factors above interact with each other. For example, someone may have a very high or low level of a given brain chemical because of genetic factors or because he/she has recently experienced a severe life event. Another example concerns the impact of cultural expectations on eating disorders. This is not the *only* factor causing eating disorders, because the overwhelming majority of women in Western societies don't suffer from eating disorders. Eating disorders occur in individuals exposed to cultural expectations of thinness *and* who are vulnerable (e.g., because of genetic factors).

In what follows, we will consider the factors responsible for various mental disorders. First, we consider major depressive disorder, which is by far the most common of the mood disorders. Next we consider some of the main anxiety disorders, which are also found in millions of individuals. Finally, we consider schizophrenia, one of the most serious mental disorders.

MOOD DISORDERS: DEPRESSION

Various mood disorders are identified in DSM-IV. Of particular importance is major depressive disorder, which affects literally millions of people around the world. Less common but more severe is a mood disorder known as bipolar disorder (see figure on the right). This is a condition involving depressive and manic (elated) episodes. Over 90% of those suffering from clinical depression have major depressive disorder rather than bipolar disorder, and so we focus on that condition. The diagnosis of **major depressive disorder** requires that five of the following symptoms occur nearly every day for 2 weeks: sad, depressed mood; loss of interest and pleasure in usual activities; difficulties in sleeping; changes in activity level; weight loss or gain; loss of energy and tiredness; negative self-concept, self-blame, and self-reproach; problems with concentration; recurring thoughts of suicide or death. About 10% of men and 20% of women become clinically depressed at some time in their lives.

Factors involved

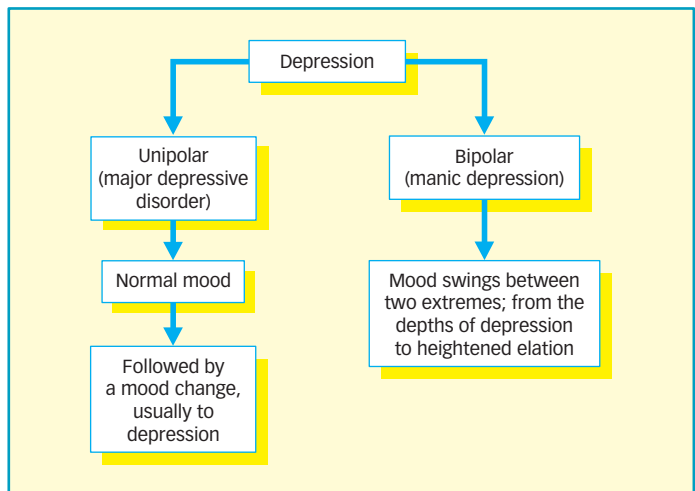
We start by considering the role of genetic factors in depression as assessed by twin studies. The key measure is the **concordance rate**: this is the likelihood that, if one twin has the disorder, the other twin also has it. If genetic factors are important, the concordance rate should be higher in identical than in fraternal twins. That is what

Case Study: Major Depressive Disorder

Chrissie seemed a successful young woman, with a good job and social life, but when a new boss started at work with a blame culture approach she rapidly felt unable to cope, worthless, apathetic, and full of self-doubts. She became exhausted, and her GP diagnosed her as having major depressive disorder and she was off work for 6 months.

As a child, Chrissie had been criticized all the time by her parents and felt she could never do anything right, not even hold her knife and fork acceptably nor eat neatly. It seems that the heavy, negative manner of her new boss had caused the memories from childhood of helplessness and failure to resurface, and the associated emotions had led to the persistent very low mood.

Treatment from a sympathetic GP and a psychotherapist who used CBT (cognitive-behavior therapy) reinforced with clinical hypnotherapy helped her recover.



Key Terms

Major depressive disorder:

a disorder characterized by symptoms such as sad, depressed mood, tiredness, and loss of interest in various activities.

Concordance rate:

as applied to twins, the probability that one twin has a given disorder given that the co-twin has it.

has been found. Allen (1976) reviewed several twin studies, and reported that the concordance rate was 40% for identical twins compared to only 11% for fraternal twins. McGuffin, Katz, Watkins, and Rutherford (1996) found that the concordance rate for major depressive disorder was 46% for identical twins and 20% for fraternal twins.

Additional evidence that genetic factors are important comes from adoption studies. Wender et al. (1986) considered adopted children who later developed major depression. The biological parents of these individuals were *eight* times as likely to have suffered from clinical depression as their adoptive parents.

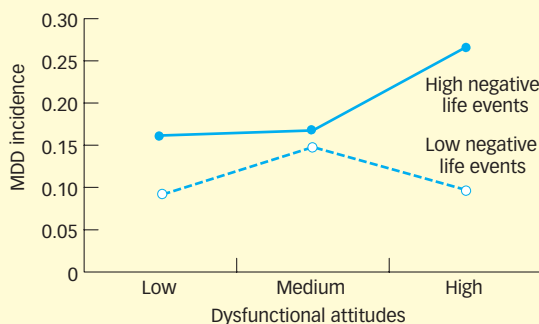
Depressed patients may have abnormal levels of various neurotransmitters or other substances. Some evidence suggests that patients with major depressive disorder often have low levels of norepinephrine and serotonin (see Comer, 2007). However, Thase et al. (2002) found that depressed patients (especially those with severe depression) had *increased* levels of norepinephrine. The reality is complex. Rampello, Nicoletti, and Nicoletti (2000) reported that patients with major depressive disorder have an *imbalance* in the activity of several neurotransmitters including norepinephrine, serotonin, dopamine, and acetylcholine. In addition, patients with major depressive disorder often have high levels of the hormone cortisol, which is present in large amounts when individuals are stressed (Thase et al., 2002).

Evidence that life events and cognitive factors can both be involved in the development of major depressive disorder was reported by Lewinsohn, Joiner, and Rohde (2001). They measured dysfunctional attitudes (e.g., “My life is wasted unless I am a success”; “I should be happy all the time”) in adolescents not having a major depressive disorder at the outset of the study. One year later, Lewinsohn et al. assessed the negative life events experienced by the participants over the 12-month period. Those who experienced many negative life events had an increased likelihood

of developing a major depressive disorder only if they were initially high in dysfunctional attitudes (see the figure on the left).

Life events often play a role in triggering depression. For example, Brown (1989) reviewed the findings from several studies. On average, about 55% of depressed patients had at least one severe life event in the months before onset, compared to only about 17% of controls. However, Brown and Harris (1978) found that the impact of life events was greatly influenced by whether the individual had an intimate friend. They focused on women who had experienced at least one very stressful life event in the 8 months before interview. Of those women who didn’t have an intimate friend, 37% became depressed. In contrast, only 10% of women having a very close friend became depressed.

Probability of developing major depressive disorder (MDD) as a function of number of life events (high vs. low) and dysfunctional attitudes (low, medium, or high)



ANXIETY

There are several anxiety disorders, and we focus on four of them. Note that anxious patients often suffer from two or more different anxiety disorders; this is known as **comorbidity**. For example, Barlow, Di Nardo, Vermilyea, and Blanchard (1986) found comorbidity in 66% of anxious patients presenting at their clinic. Hettema et al. (2005) identified a major reason for comorbidity based on a twin study focusing on the

underlying genetic risk factors for the anxiety disorders. Similar genetic influences were present for generalized anxiety disorder (persistent worry and anxiety in several areas of life), panic disorder (described below), and agoraphobia (described below).

Panic disorder is an anxiety disorder in which the individual experiences repeated panic attacks. According to DSM-IV, a panic attack involves intense fear or

Key Terms

Comorbidity:

presence in the same individual of two or more mental disorders at the same time.

Panic disorder:

a disorder in which an individual suffers from panic attacks and is very concerned about having further attacks.

discomfort, with four or more bodily symptoms suddenly appearing (e.g., palpitations, shortness of breath, accelerated heart rate, feeling of choking, sweating, chest pain, feeling dizzy, and fear of dying). Many patients with panic disorder also suffer from agoraphobia. **Agoraphobia** is a condition in which the individual is frightened to leave the house, because he/she is very concerned about having a panic attack in a public place. In DSM-IV, the term “panic disorder with agoraphobia” describes such individuals.

Posttraumatic stress disorder is triggered by a specific distressing event such as rape, war, or natural disaster. According to DSM-IV, this condition is associated with three main symptoms:

1. *Re-experiencing the event*: The event is often recalled and nightmares about it are common. Stimuli triggering memories of the traumatic event cause intense emotional upset.
2. *Avoidance of stimuli associated with the event or reduced responsiveness to such stimuli*: The individual tries to avoid trauma-related stimuli or thoughts. There is fluctuation between re-experiencing the traumatic event and a numbing of response to stimuli associated with the event.
3. *Increased arousal*: There are problems with sleep, difficulties with concentration, and an increased startle response.



Case Study: Sarah—A Case of Agoraphobia

Sarah, a woman in her mid-thirties, was shopping for bargains in a crowded department store during the January sales. Without warning and without knowing why, she suddenly felt anxious and dizzy. She worried that she was about to faint or have a heart attack. She dropped her shopping and rushed straight home. As she neared home, she noticed that her feelings of panic lessened.

A few days later she decided to go shopping again. On entering the store, she felt herself becoming increasingly anxious. After a few minutes she had become so anxious that a shop keeper asked her if she was OK and took her to a first aid room. Once there her feelings of panic became worse and she became particularly embarrassed at all the attention she was attracting.

After this she avoided going to the large store again. She even started to worry when going into smaller shops because she thought she might have another panic attack, and this worry turned into intense anxiety. Eventually she stopped shopping altogether, asking her husband to do it for her.

Over the next few months, Sarah found that she had panic attacks in more and more places. The typical pattern was that she became progressively more anxious the further away from her house she got. She tried to avoid the places where she might have a panic attack, but as the months passed, she found that this restricted her activities. Some days she found it impossible to leave the house at all. She felt that her marriage was becoming strained and that her husband resented her dependence on him.

Key Terms

Agoraphobia:

a disorder in which there is fear of public places from which it might be difficult to escape in the event of a panic attack.

Posttraumatic stress disorder:

a disorder triggered by a very distressing event and involving re-experiencing of the event, avoidance of stimuli associated with the event, and increased arousal.

Clearly Sarah's behavior was abnormal, in many of the ways described in the text. It was statistically infrequent and socially deviant. It interfered with her ability to function adequately, both from her own point of view and of her husband. She did not have many of the signs of mental healthiness.

Adapted from Stirling and Hellewell (1999).

Social phobia involves extreme concern about one's own behavior and the reactions of others. About 70% of sufferers are female and it is mostly found among young people in their late teens and twenties. The main criteria for social phobia in DSM-IV include the following:

- Marked and persistent fear of situations in which the individual will be exposed to unfamiliar people or to scrutiny.
- Exposure to the feared social situation nearly always produces considerable anxiety.
- The individual recognizes that the fear experienced is excessive.
- The feared situations are avoided or responded to with great anxiety.
- The phobic reactions interfere significantly with the individual's working or social life, or there is marked distress.

Case Study: A Social Phobic

Jim had always been shy. He had married a schoolfriend who was lively and outgoing, and she had managed their social contacts and social life for them both. Jim had coped with working in a small music store, but when this was taken over by a large chain he had to make phone calls to customers and speak to the public and this had caused such anxiety that social phobia was diagnosed. He had become too anxious to be able to do his job and felt exhausted and defeated and a failure. He lost any sense of

self-efficacy and was virtually housebound by the time he sought help.

Cognitive-behavioral therapy helped Jim to confront his phobia, see that his fears were groundless, and construct and practice strategies for managing his natural anxiety successfully.

Richards, T.A. (2008). A case study: Jim. <http://www.anxietynetwork.com/spcase.html>

Specific phobia involves strong and irrational fear of some specific object or situation. Specific phobias include fear of snakes and spiders, but can extend to fear of the number 13, fear of knees, and fear of being infested with worms.

Factors involved

What factors trigger panic disorder? Genetic factors play a role. Kendler et al. (1993) found that the concordance rate was 24% for identical twins compared to 11% for fraternal twins. Both figures are much higher than the 2% found in the population as a whole. According to the cognitive approach (e.g., Clark, 1986), patients with panic disorder greatly exaggerate the threateningness of internal stimuli (e.g., fast heart rate). This may sometimes happen because patients previously had a respiratory disease leading them to be especially sensitive about their internal state. Verburg, Griez, Meijer, and Pols (1995) found that 43% of their panic disorder patients had suffered from at least one respiratory disease during their lives compared to only 16% among patients with other anxiety disorders. Finally, life events can help to trigger panic disorder. Barrett (1979) found that panic patients reported significantly more undesirable life events in the 6 months

Key Terms

Social phobia:

a disorder in which the individual has excessive fear of many social situations and will often avoid them.

Specific phobia:

a strong and irrational fear of a given object (e.g., spider; snake) or situation (e.g., enclosed space).

before the onset of their disorder than did controls over the same time period.

Posttraumatic stress disorder (PTSD) is always triggered by some distressing event, and is especially likely to occur if the event is life threatening. However, when many people are exposed to the same traumatic event (e.g., a sinking ship), only some of them develop posttraumatic stress disorder. Thus, factors in addition to the traumatic event itself need to be taken into account. Skre et al. (1993) found that the concordance rate for posttraumatic stress disorder was greater in identical twins than in fraternal ones. Foy, Resnick, Sippelle, and Carroll (1987) found that genetic factors interacted with the severity of the traumatic event in soldiers. A low level of combat exposure was much more likely to lead to posttraumatic stress disorder in men having family members with other disorders. However, a high level of combat exposure led to posttraumatic stress disorder in two-thirds of soldiers regardless of whether family members had other disorders.

Social phobia has a genetic component. Kendler, Karkowski, and Prescott (1999) carried out a study on a large sample of female identical and fraternal twins. The concordance rate was significantly higher for identical twins than for fraternal twins. Personality (individual differences in which depend on genetic factors) seems to be involved in the development of social phobia. Stemmerger, Turner, and Beidel (1995) found that most social phobics are very introverted. Very introverted people tend to avoid social situations, and so introversion may be a risk factor for social phobia. Social factors are also involved. Arrindell et al. (1989) found that social phobics recalled their parents being rejecting, lacking in emotional warmth, or overprotective. Bruch and Heimberg (1994) found that adult social phobics said their parents overemphasized the opinions of others, failed to stress the importance of family sociability, and isolated their family from others.



PTSD can be triggered by a life-threatening event. Here, an Indian woman weeps near the debris of her damaged house hit by the 2004 Asian tsunami, the final death toll of which exceeded 225,000.

SCHIZOPHRENIA

Schizophrenia is a very serious condition. About 1% of the population of the UK suffer from schizophrenia during their lives, and the figure is similar in most other countries. The symptoms of schizophrenia vary somewhat, but typically include problems with attention, thinking, social relationships, motivation, and emotion.

Case Study: A Schizophrenic Disorder

A young man of 19 (WG) was admitted to the psychiatric services on the grounds of a dramatic change in character. His parents described him as always being extremely shy with no close friends, but in the last few months he had gone from being an average-performing student to failing his studies and leaving college. Having excelled in nonteam sports such as swimming and athletics, he was now taking no exercise at all. WG had seldom mentioned health matters, but now complained of problems with his head and chest. After being admitted, WG spent most of his time staring out of the window, and uncharacteristically not taking care over his appearance. Staff found it difficult to converse with him and he offered no information about himself, making an ordinary diagnostic interview impossible. WG would usually answer direct questions, but in a flat

emotionless tone. Sometimes his answers were not even connected to the question, and staff would find themselves wondering what the conversation had been about. There were also occasions when there was a complete mismatch between WG's emotional expression and the words he spoke. For example, he giggled continuously when speaking about a serious illness that had left his mother bedridden. On one occasion, WG became very agitated and spoke of "electrical sensations" in his brain. At other times he spoke of being influenced by a force outside himself, which took the form of a voice urging him to commit acts of violence against his parents. He claimed that the voice repeated the command "You'll have to do it."

Adapted from Hofling, 1974.

According to DSM-IV, the criteria for schizophrenia include:

1. Two or more of the following symptoms: delusions, hallucinations, disorganized speech, grossly disorganized or catatonic (rigid) behavior, and negative symptoms (lack of emotion, lack of motivation, speaking very little or uninformatively).
2. Continuous signs of disturbance for at least 6 months.
3. Social and/or occupational dysfunction or poor functioning.

Factors involved

Genetic factors play a major role in the development of schizophrenia. Gottesman (1991) summarized the findings from about 40 twin studies. The concordance rate was 48% for identical or monozygotic twins compared to only 17% for fraternal or dizygotic twins. However, genetic factors are not all-important. The finding that 50% of identical twins whose co-twin develops schizophrenia do *not* develop the disorder themselves indicates that environmental factors also play a key role in the causation of schizophrenia.

There is dispute as to whether genetic factors are as important in the development of schizophrenia as suggested by Gottesman (1991). For example, Joseph (2003) considered the nine most recent studies on schizophrenia. The concordance rate was 22.4% for identical twins and 4.6% for fraternal twins. These figures are much lower than those reported by Gottesman (1991). However, bear in mind that the lifetime risk of schizophrenia in the general population is only 1%. That means that you have a substantially increased chance of schizophrenia if you have a twin (especially identical) with the condition.

Wahlberg et al. (1997) showed how genetic factors interact with environmental ones. They studied adopted children whose biological mothers had schizophrenia. These children showed high levels of thought disorder if raised by families who communicated in unclear and confusing ways. In contrast, children with schizophrenic mothers had very good psychological health if raised by families who communicated clearly.

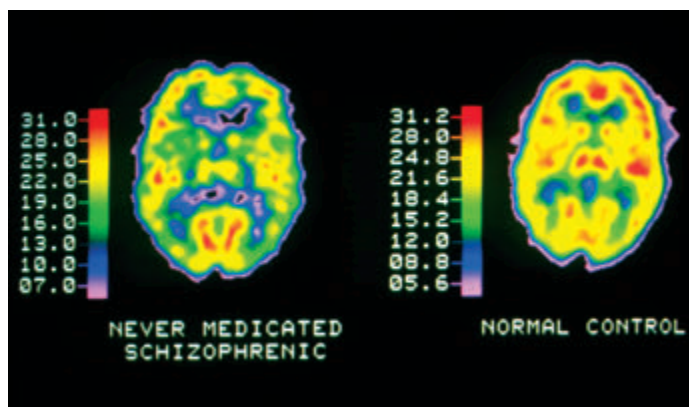
Schizophrenia is generally diagnosed when someone is in their late teens or early twenties. However, the early signs of schizophrenia can be detected in early childhood—perhaps because of the impact of genetic factors. For example, Walker et al. (1994) studied family home movies taken during the early childhood of children who did or didn't develop schizophrenia subsequently. At a very early age (sometimes as early as 2), the children who subsequently developed schizophrenia had more motor abnormalities, more negative emotional expressions, and more odd hand movements.

Social factors are emphasized by the social causation hypothesis, which was designed to explain why schizophrenics tend to belong to the lower social classes. According to this hypothesis, members of the lower social classes have more stressful lives than middle-class people, and this makes them more vulnerable to

schizophrenia. The key issue here is whether belonging to the lower social classes makes individuals likely to develop schizophrenia, or whether developing schizophrenia leads to reduced social status. Turner and Wagonfeld (1967) reported some support for the social causation hypothesis in their finding that the fathers of schizophrenics tend to belong to the lower social classes.

Stressful life events sometimes help to trigger the onset of schizophrenia. Day et al. (1987) found in several countries that schizophrenics tended to have experienced a high number of stressful life events in the few weeks before the onset of schizophrenia.

Finally, Frith (1992) argued that cognitive factors may be involved in the development of schizophrenia. He claimed that positive symptoms of schizophrenia (e.g., delusions of control; auditory hallucinations) may occur



PET scans of schizophrenic (left) and normal brains (right). The darkest areas show low levels of activity.

because schizophrenics have problems with self-monitoring. Patients fail to keep track of their own intentions, as a consequence of which they mistakenly experience their own thoughts as alien. McGuire et al. (1996) reported support for this theory. They found that schizophrenics suffering from hallucinations had reduced activity in those parts of the brain involved in monitoring inner speech.

Chapter Summary

What is abnormality?

- Decisions about what is abnormal depend in part on the prevailing social norms and values.
- Four features typically associated with abnormality are deviance, distress, dysfunction, and danger. However, these criteria are vague and often difficult to apply in practice.

Classification systems

- The two main classification systems are DSM and ICD.
- DSM consists of five axes. It refers to culture-bound syndromes but de-emphasizes cross-cultural differences.
- Most classification systems make use of categories. However, dimensional systems that avoid rigid categorization are more realistic.

Factors causing mental disorder

- According to the diathesis–stress model, the occurrence of mental disorder depends on individual vulnerability and stressful events.
- Factors involved in the development of major depressive disorder include genetic influences, life events, and dysfunctional attitudes.
- Genetic factors play a role in triggering the main anxiety disorders. Previous history of respiratory disease is associated with panic disorder, severity of the traumatic event with posttraumatic stress disorder, and an introverted personality and rejecting or overprotective parents with social phobia.
- Genetic factors play a major role in the development of schizophrenia. Social factors, stressful life events, and problems with self-monitoring are also involved.

Further Reading

- Bennett, P. (2003). *Abnormal and clinical psychology: An introductory textbook* (2nd ed.). Maidenhead, UK: Open University Press. There is accessible coverage of the main issues within abnormal psychology in this readable book.
- Comer, R.J. (2007). *Abnormal psychology* (6th ed.). New York: Worth. Most of the important issues are discussed in a thorough and accessible way in this up-to-date textbook.
- Davison, G.C., Neale, J.M., & Kring, A.M. (2004). *Abnormal psychology* (9th ed.). New York: Worth. This well-established textbook has readable accounts of the whole of abnormal psychology.
- Durand, V.M., & Barlow, D.H. (2006). *Essentials of abnormal psychology* (4th ed.). New York: Thomson/Wadsworth. As the title implies, this textbook by leading clinical psychologists focuses on central issues within abnormal psychology.

chapter 22

Contents

Drug therapy	523
Psychodynamic therapy	525
Behavior therapy	527
Cognitive and cognitive-behavior therapy	529
Effectiveness of therapy	532
How effective are different forms of therapy?	533

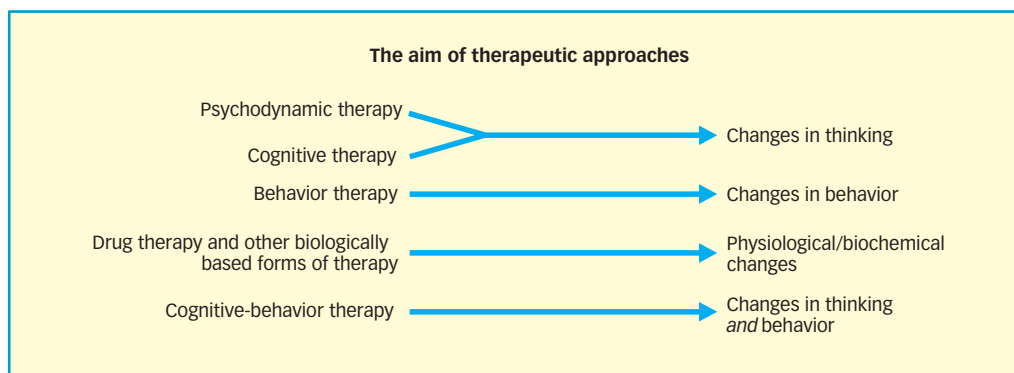
Therapeutic approaches

22

Individuals with mental disorders exhibit a wide range of symptoms. There may be problems associated with thinking and the mind (e.g., the hallucinations of the schizophrenic), with behavior (e.g., the avoidance behavior of the phobic), or with physiological and bodily processes (e.g., the highly activated physiological system of someone with posttraumatic stress disorder). However, note that thinking, behavior, and physiological processes are all highly *interdependent*, meaning that they all affect each other.

Therapeutic approaches could focus on producing changes in thinking, in behavior, or in physiological functioning. That is precisely what has happened. The psychodynamic approach was designed to change thinking, and the same is true of cognitive therapy. Behavior therapy emphasizes the importance of changing behavior. Drug therapy and other biologically based forms of therapy focus on physiological and biochemical changes. Finally, cognitive-behavior therapy falls somewhere between behavior therapy and psychodynamic therapy, in that it attempts to produce changes in clients' thought processes *and* in their behavior. In spite of the differing emphases across therapies, most therapists assume that their preferred form of therapy will have beneficial effects in thinking, behavior, and physiological functioning.

When considering various forms of therapy, we must avoid making the treatment etiology fallacy (MacLeod, 1998). This is the mistaken notion that the success of a given form of treatment reveals the cause of the disorder. For example, aspirin is an effective cure for headache. However, that doesn't mean that a lack of aspirin causes headaches!



DRUG THERAPY

We will consider drug therapy for depression, anxiety, and schizophrenia in turn. As we saw earlier, patients with major depressive disorder often have low levels of the neurotransmitters serotonin and norepinephrine. Tricyclic antidepressant drugs appear to increase the activity of those neurotransmitters, and generally reduce the symptoms of depression (Elkin, 1994). However, there can be a high relapse rate



Prozac is an antidepressant drug which inhibits the uptake of serotonin, a neurotransmitter produced in the brain.

unless drug therapy is continued over a long period of time. The most common drugs used to treat depression are the selective serotonin reuptake inhibitors (SSRIs), of which Prozac is the best-known. They increase serotonin activity and are as effective as the tricyclics (Hirschfeld, 1999). They have the advantage over the tricyclics that it is harder to overdose on SSRIs.

The benzodiazepines (e.g., Librium and Valium) are often used to treat anxiety disorders. In 2003, over 13 million prescriptions for benzodiazepines were written in the UK (Garfield, 2003). The benzodiazepines are reasonably effective in the treatment of generalized anxiety disorder (Rickels, DeMartinis, & Aufdembrinke, 2000), and have also been used to treat social phobia. Mitte (2005) found in a meta-analysis of 65 studies that drug-based approaches to the treatment of generalized anxiety disorder are mostly effective, and

are comparably effective to cognitive-behavior therapy. In spite of these successes, there are several unwanted effects of benzodiazepines:

1. Anxious symptoms often return when patients stop taking the drugs.
2. There are various side effects (e.g., lack of coordination; poor concentration; memory loss).
3. There can be physical dependence, with patients finding it difficult to manage without drugs.

Tricyclic drugs (generally used to treat major depression) have been used successfully to treat panic disorder. Barlow, Gorman, Shear, and Woods (2000) found that tricyclic drugs were as effective as cognitive-behavior therapy in reducing the symptoms of panic disorder, but were less effective at 6-month follow-up.

Schizophrenia is increasingly treated with atypical antipsychotic drugs (e.g., Clozaril; Risperdal; Zyprexa). These drugs have fewer side effects than the previous neuroleptic drugs used to treat schizophrenia. However, patients taking Clozaril have a 1–2% risk of developing a life-threatening condition involving a substantial reduction in white blood cells. The good news is that atypical drugs benefit 85% of schizophrenic patients compared to 65% given neuroleptic drugs (Awad & Voruganti, 1999).

Evaluation

- + Drug therapy often produces rapid beneficial effects when used to treat depression, anxiety, and schizophrenia.
- + Schizophrenia is notoriously difficult to treat, and drug therapy is probably more effective at reducing its symptoms than any other approach.
- Relapse is more common after drug therapy than after other types of therapy, probably because drugs don't deal directly with the problems underlying any given mental disorder.
- It is often difficult to know precisely *why* any given drug is effective in the treatment of a particular disorder because of our limited understanding of the underlying biochemical factors.
- The drop-out rate is often rather high when drug therapy is used.
- Nearly all drugs have unwanted side effects.

PSYCHODYNAMIC THERAPY

The first form of psychodynamic therapy was psychoanalysis, developed by Sigmund Freud early in the twentieth century. According to Freud, neuroses such as the anxiety disorders occur because of conflicts among the three parts of the mind: the ego (rational mind), the id (sexual and other instincts), and the superego (conscience). These conflicts cause the ego to use various defense mechanisms to defend itself. The key defense mechanism is repression. **Repression** involves forcing painful, threatening, or unacceptable thoughts and memories out of consciousness into the unconscious mind. Repressed memories mostly refer to childhood and to the conflicts between the instinctive (e.g., sexual) motives of the child and the restraints imposed by his/her parents. Evidence for repression is discussed in Chapter 8.

According to Freud, adults experiencing great personal problems tend to show regression. **Regression** involves going backwards through the stages of psychosexual development they went through in childhood (see Chapter 12). Children often fixate or spend an unusually long time at a given stage of psychosexual development if it was associated with conflicts or with excessive gratification. Regression typically occurs back to a stage at which the person had previously fixated.

Freud argued that the way to cure neurosis was to allow the client to gain access to his/her repressed ideas and conflicts and to face up to whatever emerged from the unconscious. He insisted that the client should focus on the feelings associated with the repressed ideas, and shouldn't simply regard them unemotionally in order to gain insight. Insight "involves a conscious awareness of some of the wishes, defenses, and compromises . . . that have interacted to produce emotional conflict or deficits in psychological development" (Kivlighan, Multon, & Patton, 2000, p. 50).

How can we uncover repressed memories and allow the client to gain insight? One method used by Freud was **free association**, in which the client said the first thing coming into his/her mind. Free association is ineffective if the client shows resistance and is reluctant to say what he/she is thinking. Nevertheless, the presence of resistance (revealed by long pauses) suggests that the client is getting close to an important repressed idea, and that further probing by the therapist is needed.

Another method is the analysis of dreams, which Freud described as "the via regia [royal road] to the unconscious." Freud argued that the mind has a censor that keeps repressed material out of conscious awareness. This censor is less vigilant (it nods off?) during sleep, and so repressed ideas from the unconscious are more likely to appear in dreams than in waking thought. However, these repressed ideas usually emerge in disguised form because of their unacceptable nature. As a result, the therapist has to work with the client to decide on the true meaning of each dream (see Chapter 5 for more on Freud's theory of dreaming).

Progress in therapy depends partly on **transference**. This involves the client transferring onto the therapist the powerful emotional reactions previously directed at his/her own parents or other highly significant individuals.

FINDINGS

There is support for the notion that at least some adult mental disorders have their origins in childhood. For example, Kendler, Pedersen, Farahmand, and Persson (1996) studied adult female twins who had experienced parental loss through separation in childhood. These twins tended to suffer from depression and alcoholism in adult



Freud asked his patients to lie on a couch during psychoanalysis. This is a photograph of his couch in his London house, which is now the Freud Museum.

Key Terms

Repression: motivated forgetting of traumatic or other very threatening events.

Regression: in Freudian theory, returning to an earlier stage of psychosexual development when someone is highly stressed.

Free association: a technique used in psychoanalysis in which the client says whatever comes immediately to mind.

Transference: in Freudian theory, the transfer of the patient's strong feelings for one person (e.g., a parent) onto the therapist.

Case Study: Anna O

Freud's theory was largely based on the observations he made during consultations with patients. He suggested that his work was similar to that of an archaeologist, who digs away layers of earth before uncovering what he or she was seeking. In a similar way, the psychoanalyst seeks to dig down to the unconscious and discover the key to the individual's personality dynamic.

Anna O was a girl of twenty-one, of a high degree of intelligence. Her illness first appeared while she was caring for her father whom she tenderly loved, during the severe illness which led to his death. The patient had a severe paralysis of both right extremities, disturbance of eye-movements, and intense nausea when she attempted to take nourishment, and at "one time for several weeks a loss of the power to drink, in spite of tormenting thirst." She occasionally became confused or delirious and mumbled several words to herself. If these same words were later repeated to her, when she was in a hypnotic stage, she engaged in deeply sad, often poetically beautiful, day dreams, we might call them, which commonly took as their starting point the situation of a girl beside the sick-bed of her father. The patient jokingly called this treatment "chimney sweeping."

Dr Breuer [Freud's colleague] soon hit upon the fact that through such cleansing of the soul more could be accomplished than a temporary removal of the constantly recurring mental "clouds."

During one session, the patient recalled an occasion when she was with her governess, and how "that lady's little dog, that she abhorred, had drunk out of a glass. Out of respect for the conventions the patient had remained silent, but now under hypnosis she gave energetic expression to

her restrained anger," and then drank a large quantity of water without trouble, and woke from hypnosis with the glass at her lips. "The symptom thereupon vanished permanently."

"Permit me to dwell for a moment on this experience. No one had ever cured an hysterical symptom by such means before, or had come so near [to] understanding its cause. This would be a pregnant discovery if the expectation could be confirmed that still other, perhaps the majority of symptoms, originated in this way and could be removed by the same method."

Such was indeed the case, almost all the symptoms originated in exactly this way, as we were to discover. The patient's illness originated at the time when she was caring for her sick father, "and her symptoms could only be regarded as memory symbols of his sickness and death."

"While she was seated at her father's sick bed, she was careful to betray nothing of her anxiety and her painful depression to the patient. When, later, she reproduced the same scene before the physician, the emotion which she had suppressed on the occurrence of the scene burst out with especial strength, as though it had been pent up all along."

"In her normal state she was entirely ignorant of the pathogenic scenes and of their connection with her symptoms. She had forgotten those scenes. . . . When the patient was hypnotized, it was possible, after considerable difficulty, to recall those scenes to her memory, and by this means of recall the symptoms were removed."

Adapted from Freud (1910).

life. Caspi, Moffitt, Newman, and Silva (1996) found that children who were inhibited at the age of 3 tended to be depressed at the age of 21.

It is difficult to test Freud's notion that insight is of crucial importance, because the concept is rather vague. For example, Hoglend et al. (1994) found that psychiatrists showed poor agreement among themselves concerning the insight levels shown by anxious and depressed patients in therapy. However, impressive findings using the Insight Rating Scale (in which expert judges evaluate clients' reports of the counseling session just finished) were reported by Kivlighan et al. (2000). They studied 12 clients who received 20 sessions in psychoanalytic counseling. There were steady decreases in symptoms across sessions and steady increases in insight. Of key importance, clients showing an increase in insight had reduced symptoms over the following week. This suggests that (as predicted by Freud) insight plays a role in causing symptom reduction.

Eysenck (1952) reviewed several studies, and claimed that the recovery rate was 72% for clients receiving no systematic treatment against only 44% for those receiving psychoanalysis. These findings suggest that psychoanalysis is actually bad for you! However, Eysenck regarded partially recovered clients as not having recovered at all, which doesn't make much sense. When the same data were re-analyzed using more reasonable criteria to define recovery, the success rate for psychoanalysis was 83% compared to 30% for those not treated (Bergin, 1971).

There is a lack of well-controlled studies on the effectiveness of psychodynamic therapy, especially with respect to the anxiety disorders (Fonagy, Roth, & Higgitt, 2005). However, such therapy has proved useful in the treatment of major depressive disorder and schizophrenia (Jarvis, 2004).

Evaluation

- + Psychoanalysis was the first systematic form of psychological treatment for mental disorders, and has proved moderately effective (see Lambert & Bergin, 1994).
- + Psychoanalysis strongly influenced the development of later forms of psychodynamic therapy and also cognitive therapy with its emphasis on the individual's mistaken interpretations of himself/herself.
- “The concepts on which it [the psychodynamic model] are based are difficult to define and to research . . . Because processes such as id drives, ego defenses, and fixation are abstract and supposedly operate at an unconscious level, there is no way of knowing for certain if they are occurring” (Comer, 2001, p. 59).
- Freud argued that insight is needed to produce recovery, and some evidence supports that view (e.g., Kivlighan et al., 2000). However, it is generally unclear whether insight causes recovery or whether recovery is needed to produce insight.
- Much of what the client says may be influenced by suggestions implanted previously by the therapist and so may not reflect his/her genuine views.

BEHAVIOR THERAPY

Behavior therapy developed during the 1950s and 1960s, but its origins go back several decades before that. The underlying notions are that most mental disorders are caused by maladaptive learning, and that the best treatment consists of appropriate new learning. Behavior therapists believe that abnormal behavior develops through conditioning (see Chapter 7). There are two main forms of conditioning: (1) classical conditioning involving learning by association; and (2) operant conditioning involving learning by reinforcement or reward. Behavior therapists argue that classical and operant conditioning can change unwanted behavior into a more desirable pattern of behavior. Some experts limit the term “behavior therapy” to treatment using classical conditioning, and use the term “behavior modification” to refer to treatment based on operant conditioning.

Behavior therapy has an emphasis on *current* problems and behavior, and on removing any symptoms the client finds troublesome. This contrasts with psychoanalysis, where the focus is on uncovering unresolved conflicts from childhood. In principle, behavior therapy has the advantage that it is based on scientific evidence acquired from laboratory studies of conditioning.

Here are three forms of behavior therapy:

1. **Exposure therapy.** **Exposure therapy** involves putting the client in feared situations (e.g., social situations for social phobics), with the feared situation becoming gradually more intense over time. It is used extensively in the treatment of phobias or extreme fears. Exposure therapy has proved successful in the treatment of most anxiety disorders (see Eysenck, 1997). It is often

Key Term

Exposure therapy: a form of treatment in which clients are repeatedly exposed to stimuli or situations they greatly fear.



regarded as the “gold standard” against which other forms of therapy should be compared. It is assumed that phobic fears are acquired by classical conditioning, with the conditioned or phobic stimulus (e.g., a snake) being associated with a painful or aversive stimulus creating fear. The repeated presentation of the conditioned stimulus in exposure therapy leads to extinction or habituation of the fear response.

2. *Aversion therapy.* In classical aversive conditioning, a neutral or positive stimulus is paired with an unpleasant or aversive stimulus (e.g., electric shock). This causes an aversive reaction (e.g., anxiety) to the neutral or positive stimulus. **Aversion therapy** is based on aversive conditioning. For example, an alcoholic is given aversive stimuli (e.g., electric shocks) while starting to drink alcohol. The intention is that alcohol will produce feelings of anxiety and so inhibit drinking. Alcoholism has also been treated by using drugs causing vomiting or impeding breathing (e.g., Antabuse). There are promising signs that drugs based on tryptophan metabolites may prove useful in treating alcoholism. They make patients feel ill if they have even a small amount of alcohol. In addition, they increase serotonin levels and so enhance patients’ sense of wellbeing if they don’t drink (see Martin & Bonner, 2005). Aversion therapy can be fairly effective in the short term, but generally not in the long term (Comer, 2007; Roth & Fonagy, 2005).
3. *Token economy.* The **token economy** involves selective positive reinforcement or reward. It is used with institutionalized patients who are given tokens (e.g., colored counters) for behaving appropriately. These tokens can be used to obtain various privileges (e.g., obtaining cigarettes). Ayllon and Azrin (1968) carried out a classic study. Female schizophrenic patients who had been hospitalized for an average of 16 years were given plastic tokens for actions such as making their beds or combing their hair. The tokens were exchanged for pleasant activities (e.g., seeing a movie). Use of a token economy increase eight-fold the number of chores the patients performed every day. Dickerson, Tenhula, and Green-Paden (2005) reviewed 13 studies of the token economy being used with schizophrenics. Beneficial effects were reported in 11 of these studies. The token economy is not used much nowadays in the treatment of schizophrenics, who are now mostly treated in the community. One reason for reduced use of the token economy is that it is very hard to exert the required level of environmental control within the community.

Evaluation

- + Behavior therapy has proved effective in treating several mental disorders, especially those with easily identifiable behavioral symptoms (e.g., avoidance responses of phobic individuals). For example, exposure therapy has many successes to its credit.
- + Behavior therapy has a scientific basis and conditioning experiences are sometimes involved in the development of mental disorders.
- The role of environmental factors in causing mental disorders is exaggerated, and that of genetic and other biological factors minimized.
- Some forms of behavior therapy are not very effective. As Roth and Fonagy (1996, p. 226) pointed out, “There is little support for techniques using chemical or electrical aversion. Its use in service settings is made even more difficult to recommend given that there is (unsurprisingly) a high rate of attrition [dropping out from therapy].” Aversion therapy also poses ethical issues because of the high levels of discomfort and stress involved.
- Token economies sometimes produce token (i.e., minimal) learning. They work because the environment is carefully structured so that only good behavior is consistently rewarded. Patients find it hard to transfer what they have learned to the much less structured environment provided by the real world.

Key Terms

Aversion therapy:

a form of treatment in which an aversive stimulus is paired with a positive one (e.g., sight of alcohol) to inhibit the response to the positive stimulus (e.g., drinking).

Token economy:

use of **operant conditioning** to change the behavior of mental patients by selective positive reinforcement or reward.

COGNITIVE AND COGNITIVE-BEHAVIOR THERAPY

The cognitive approach to therapy was developed mainly by Albert Ellis (1962) and Aaron Beck (1967). This approach is based in part on the psychodynamic model. This is especially so with its emphasis on the important role played by mental processes in the development and maintenance of mental disorder. However, as we will see, there are many major differences between the two approaches.

The central notion in the cognitive approach is that individuals suffering from mental disorders have distorted and irrational thoughts and beliefs. In practice, this approach has mainly been applied to major depressive disorder and to the anxiety disorders. Warren and Zgourides (1991) discussed the kinds of distorted beliefs much more common among individuals with anxiety disorders than among healthy individuals. Examples include, “I *must* perform well and/or win the approval of others, or else it’s awful,” and “My life conditions *must* give me the things I want easily and with little frustration . . . or else life is unbearable.”

Beck (1976) argued that many of the cognitive distortions of depressed patients center around the **cognitive triad** (see the figure on the right). This consists of negative thoughts that depressed individuals have about themselves, about the world, and about the future. Depressed patients typically regard themselves as helpless, worthless, and inadequate. They interpret events in the world in an unrealistically negative and defeatist way. The final part of the cognitive triad involves depressed individuals seeing the future as totally hopeless, because their worthlessness will prevent any improvement occurring in their situation.

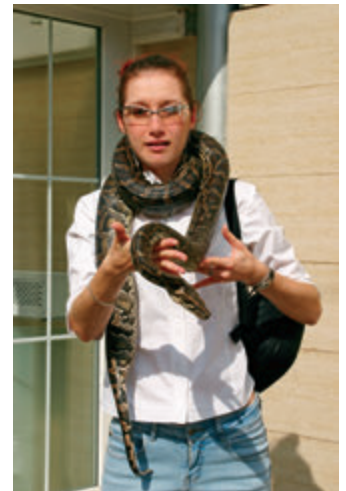
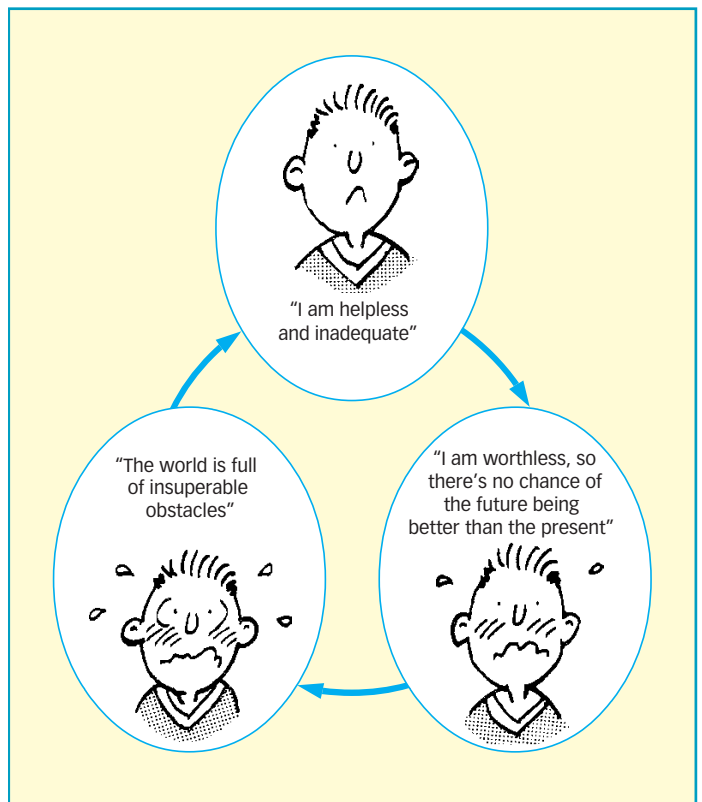
According to Beck, two maladaptive forms of thinking in depressed patients are negative automatic thoughts and overgeneralization. Negative automatic thoughts (e.g., “I always make a mess of things”) are triggered effortlessly when depressed individuals experience failure. Overgeneralization involves drawing very general negative conclusions from specific evidence. For example, failing to obtain one particular job is interpreted as meaning that the depressed person will never find a job again.

What about the anxiety disorders? It is assumed that anxious clients overestimate the threateningness of certain external or internal stimuli. For example, spider phobics exaggerate the threat posed by spiders, and patients with panic disorder exaggerate the threateningness of an aroused bodily state. According to Beck, Emery, and Greenberg (1985), the level of anxiety that an individual will experience when facing a threatening event can be estimated from the following equation:

$$\frac{\text{Perceived probability of threat} \times \text{perceived cost of event}}{\text{Perceived ability to cope} \times \text{perceived “rescue factors”}}$$

Individuals with anxiety disorders exaggerate the probability of a threatening event happening and the psychological cost of that event happening. In addition, they minimize their ability to cope with such an event and doubt whether rescue factors (e.g., help from others) will be forthcoming.

Why do anxious patients maintain distorted and irrational thoughts and beliefs year after year in the face of contrary evidence from the world around them? The short answer is **safety-seeking behaviors**. These are forms of behavior designed to prevent feared events from happening and thus reduce anxiety. For example, social phobics believe mistakenly that safety-seeking behaviors such as avoiding eye contact, talking very little, and not



This girl is happy to handle this snake, but a snake phobic, according to Beck, would overestimate the danger the snake posed.

Key Terms

Cognitive triad:

the depressed person's negative views of the self, the world, and the future.

Safety-seeking behaviors:

actions taken by anxious patients to prevent feared consequences and to reduce their anxiety level.

Common irrational beliefs

- I must be loved or liked by every significant person I meet
- I must be perfect if I am worthwhile
- The world will end if things are not as I want them to be
- My unhappiness is not under my own control and I am powerless to change things
- I should worry about bad things happening
- It is better to put off dealing with anything unpleasant than to face it
- I need a stronger person to depend on
- My problems are the result of my past

talking about themselves prevent social catastrophe (Clark & Wells, 1995). In similar fashion, some patients with panic disorder believe that sitting quietly and controlling their breathing reduces the severity of panic attacks and wards off heart attacks.

There is convincing evidence that anxious and depressed patients have irrational beliefs. For example, Newmark, Frerking, Cook, and Newmark (1973) found that 65% of anxious patients (but only 2% of healthy controls) agreed with the statement, “It is essential that one be loved or approved of by virtually everyone in his community.” However, the crucial issue is whether these beliefs help to produce clinical anxiety and depression (as

assumed by cognitive therapists) or whether these irrational beliefs are a by-product of becoming clinically anxious or depressed. Some evidence that irrational beliefs can be involved in triggering clinical depression was discussed in Chapter 21. Lewinsohn, Joiner, and Rohde (2001) found that severe life events were followed by major depressive disorder mainly in those individuals who were previously found to have irrational beliefs.

What is involved in cognitive therapy? According to Beck and Weishaar (1989, p. 308), “Cognitive therapy consists of highly specific learning experiences designed to teach patients (1) to monitor their negative, automatic thoughts (cognitions); (2) to recognize the connections between cognitions, affect [emotion], and behavior; (3) to examine the evidence for and against distorted automatic thoughts; (4) to substitute more reality-oriented interpretations for their biased cognitions; and (5) to learn to identify and alter the beliefs that predispose them to distort their experiences.”

Cognitive therapy differs from psychodynamic therapy in that the focus is on the patient’s current concerns and beliefs rather than those of childhood. It also differs in that “The concept of unconscious processes is largely irrelevant to cognitive therapy” (Sacco & Beck, 1985, p. 5).

Consider as an example cognitive therapy for social phobia, in which there is extreme fear of social situations. Social phobics’ belief that their social behavior is very inadequate

Case Study: Social Phobia

Dinesh is a 25-year-old software engineer. He was an award-winning student at engineering college. After graduating he got a job and within a short span of time impressed his superiors with his quality of work.

In spite of his achievements, Dinesh faced social problems. In social situations he felt that everybody was watching him and that he might do something embarrassing. At official meetings, he was tense and did not express his views. The most difficult task was attending parties. He was extremely tense, at times trembling, giddy with fear that others might ridicule him. Then he began to stay away.

Dinesh was suffering from social phobia, the irrational fear of social situations. Sufferers are scared and anxious in social situations. The anxiety is because they feel that something awful will happen to them, for example, they may behave in an embarrassing way and others will ridicule them. This results in sweating, trembling, palpitations, and feeling giddy. Even though the individual is aware that such behavior is silly, they are not able to overcome it. They start avoiding social situations and this results in difficulties in

carrying out their work and personal activities. This deteriorated interaction with others and personal distress results in a lower level of functioning and low self-esteem.

For those afflicted with the disorder, both medical and nonmedical treatments are available. In terms of medication, drugs like fluoxetine and recent arrivals like moclobemide are useful. But drugs take effect only after a few weeks, and do not provide a permanent cure.

Psychological treatment in the form of cognitive-behavior therapy is useful. This aims at uncovering a person’s automatic negative thoughts and cognitive schema, and helps them to understand and overcome these negative thoughts, in a gradual and systematic way. Relaxation therapy is also useful in helping an individual overcome basic anxiety.

Coming back to Dinesh, he was on medication and underwent cognitive-behavior therapy. Gradually, he overcame his phobia. He is now attending parties and meetings, is able to travel, and feels good about himself.

Adapted from Anandaram (2001).

can be challenged by presenting them with video recordings of their actual social behavior (Harvey, Clark, Ehlers, & Rapee, et al., 2000). Their belief that avoiding eye contact and other safety-seeking behaviors serves to prevent social catastrophe can be challenged by instructing them not to use any of these behaviors in social situations. This has proved effective in therapy (e.g., Morgan & Raffle, 1999).

Cognitive therapists discovered that restructuring a client's thoughts and beliefs didn't always produce full recovery. For example, a social phobic might accept that there is much less to fear in social situations than they had previously believed. However, he/she might still be reluctant to put himself/herself in difficult situations. This led to the development of cognitive-behavior therapy. The basic notion is that the client needs to change his/her dysfunctional thoughts *and* his/her inappropriate behavior. Kendall and Hammen (1998) argued that four basic assumptions underlie cognitive-behavior therapy:

1. Patients typically respond on the basis of their *interpretations* of themselves and the world around them rather than what is *actually* the case.
2. Thoughts, behaviors, and feelings are all interrelated, and all influence each other. Thus, no single factor is more important than the others.
3. In order for therapeutic interventions to be successful, they must clarify and change how people think about themselves and the world around them.
4. It is important to change both the client's cognitive processes and his/her behavior. The reason is that the benefits of therapy are likely to be greater than when only cognitive processes or behavior change.

Clark (1996) provided an example of cognitive-behavior therapy involving a 40-year-old man with panic disorder. The patient tried to protect himself against having a heart attack during panic attacks by taking paracetamol and by taking deep breaths. This hypothesis was tested by the therapist and the patient alternately sprinting and jogging around a football pitch. In addition, the patient was given the homework of taking strenuous daily exercise without trying to control his breathing. The patient rapidly accepted that his problem centered on his own mistaken beliefs.

Butler et al. (2005) considered the findings from 16 meta-analyses concerned with the therapeutic value of cognitive-behavior therapy. This form of therapy was notably successful with major depressive disorder, generalized anxiety disorder, panic disorder with or without agoraphobia, social phobia, and posttraumatic stress disorder. In addition, cognitive-behavior therapy was more effective than antidepressants in the treatment of major depressive disorder, and had beneficial effects when used with schizophrenic patients.

Evaluation

- + Cognitive therapy is effective in the treatment of anxiety disorders and depression, and cognitive-behavior therapy is even more effective (Butler et al., 2005).
- + The effectiveness of exposure therapy (used in behavior therapy) is increased when patients avoid safety-seeking behaviors as predicted by cognitive-behavior therapists (e.g., Salkovskis et al., 1999).
- Cognitive-behavior therapy is less effective in treating schizophrenia than depression or anxiety.
- There is a danger of exaggerating the importance of cognitive processes. Many clients develop more rational and less distorted ways of thinking about important issues with no beneficial changes in their maladaptive behavior.

Case Study: Acrophobia: Fear of Heights

Acrophobics have an overpowering fear of heights, a phobic fear that produces intense anxiety, panic attacks, and sends heart rate and blood pressure sky high. Traditionally, therapy has been based on the usual approaches—psychodynamic, behavioral, cognitive, bio-medical, and so on. However, a novel and highly successful new technique, virtual reality (VR), has been developed in California by Lamson (2004).

Sixty participants were recruited from advertisements and were assigned randomly to either the experimental (VR) or the control (drugs) group. Each participant set themselves a target to be reached after a week's therapy, and over 90% of the VR group achieved theirs.

The VR group wore headsets that took them into a cafe with an open terrace. During the 50-minute therapy session the participant had to “walk” across the terrace and over a

plank that led to a bridge set in a hilly scene. Lamson monitored participants' blood pressure and heart rate, and found that the fear response did subside during the 50-minute session. This therapy is a development of exposure therapy, but avoids the extreme anxiety of confronting the phobic stimulus in reality. It also avoids the risks, such as side effects, of even successful drug therapy. Perhaps one of the strengths of VR therapy is that it gives the clients confidence; they have been successful in the virtual world so this encourages them to be optimistic about success in the real world. As Lamson says, “within 3 months the woman who could barely start up a step ladder was on the roof cleaning the gutters.”

Lamson, R. (2004). <http://www.newscientist.com/channel/health/mental-health/mg14219290.700.html>

EFFECTIVENESS OF THERAPY

On the face of it, it seems easy to assess the effectiveness of therapy, and to decide whether one form of therapy is more or less effective than another. We could carry out a study including three groups of clients having the same diagnosis:

- One group receives therapy A.
- One group receives therapy B.
- One group remains on a waiting list.

We could then compare the percentages of clients recovering in the three groups, which would reveal the relative effectiveness of therapies A and B. However, there are several problems in assessing therapeutic effectiveness, some of which are discussed below.

First, there are several ways of assessing recovery, and therapists differ in what they regard as suitable outcome measures. For example, a major goal of therapy for psychodynamic therapists is to resolve inner conflicts, whereas for behavior therapists it is to produce desirable changes in overt behavior. Strupp (1996) pointed out that the effectiveness of any therapy can be considered from three perspectives:

1. *Society's perspective*: The individual's ability to function in society; the individual's adherence to social norms.
2. *Client's own perspective*: His/her overall subjective wellbeing; his/her ability to function effectively socially and at work.
3. *Therapist's perspective*: The client's thinking and behavior are related to the theoretical framework underlying the therapy used by the therapist.

Not surprisingly, the extent to which a client has recovered may vary considerably depending on which perspective we adopt.

Second, there is a conflict at the heart of much research on therapeutic effectiveness. This conflict involves deciding whether to adopt a scientific approach involving a high level of control over what happens, or to adopt a more realistic approach rooted in clinical practice. This conflict is a very real one because it is seldom possible to be scientific *and* realistic at the same time. Seligman (1995) focused on this conflict in his distinction between efficacy studies and effectiveness studies. **Efficacy studies** are scientific, well-controlled clinical trials (e.g., holding constant the number of therapy sessions) focusing on the elimination of well-defined problems. In contrast, **effectiveness studies** deal with actual clinical practice with all of its scientific shortcomings, and the focus is on subjective outcome measures (e.g., quality of life).

Key Terms

Efficacy studies:

assessments of therapeutic effectiveness based on well-controlled investigations of well-defined clinical problems.

Effectiveness studies:

assessments of therapeutic effectiveness based on typical clinical practice; see **efficacy studies**.

Should researchers carry out efficacy studies or effectiveness studies? Both kinds of studies are worthwhile. Efficacy studies (often regarded as the “gold standard” or best approach) allow us to identify factors responsible for benefit to clients and thus to interpret the findings with confidence. However, it is difficult to *generalize* the findings from such controlled studies to typical clinical practice. The strengths and weaknesses of effectiveness studies are exactly the opposite. They are informative about the typical outcomes in clinical practice. However, the uncontrolled nature of such studies makes it hard to be sure about the findings reported or about the interpretations of those findings. Our confidence in any given form of therapy would clearly be greatest if it produced good outcomes in efficacy *and* effectiveness studies. The limited evidence available suggests that beneficial effects of therapy tend to be greater in efficacy studies than in effectiveness ones (Weisz et al., 1995), in part because there is more focus on adherence to detailed treatment plans in the former studies.

Third, most waiting-list clients are actively seeking advice from other people while awaiting treatment (Cross, Sheehan, & Khan, 1980). Surprisingly, clients receiving treatment seek outside help even more than waiting-list controls. Changes in treated and untreated client groups over time depend to an unknown extent on these nontherapist sources of assistance.

HOW EFFECTIVE ARE DIFFERENT FORMS OF THERAPY?

In spite of the above issues, we can work out approximately the effectiveness of any form of therapy. Increasingly, researchers make use of meta-analysis, in which the findings from numerous studies are combined to provide an overall estimate of therapeutic effectiveness. For example, Shadish, Matt, Navarro, and Phillips (2000) conducted a meta-analysis of studies in which therapy was carried out under typical clinical conditions. All major forms of therapy were effective, and there were only small differences in effectiveness among them.

The finding that all forms of therapy are of comparable effectiveness has by no means always been reported—see the Key Study below.

Key Study

Matt and Navarro (1997): A review of the effects of therapy

Matt and Navarro (1997) considered 63 meta-analyses in which different types of therapy had been compared. Altogether, these meta-analyses were based on over 3800 studies involving tens of thousands of patients. On average, 75% of patients receiving treatment improved more than the average untreated control patient. Moderate beneficial effects were found with all forms of therapy: “The general class of interventions referred to as psychotherapy . . . appear to have universally positive effects compared to no-treatment controls across a wide range of different classes of interventions, patient populations, settings, outcomes, and research designs . . . Not a single meta-analysis concluded that psychotherapeutic interventions were ineffective” (Matt & Navarro, 1997, p. 5).

In about one-third of these meta-analyses, differences were reported in the effectiveness of different therapies. In general, behavior therapy and cognitive therapy tended to be more effective than psychodynamic or client-centered therapy. However, Matt and Navarro argued that those differences were more apparent than real. Clients treated by behavior or cognitive therapy often had less serious symptoms than those treated by psychodynamic or client-centered therapy. Nevertheless, it is striking that there are remarkably few studies in which psychodynamic therapy has been shown to be significantly more effective than alternative forms of therapy (Fonagy et al., 2005).

Matt and Navarro (1997) accepted that the information obtained from meta-analyses is generally limited in some ways. For example, it seems likely that the effectiveness of any given form of psychotherapy depends on several factors such as the nature of the disorder, the seriousness of the disorder, patient characteristics, and the precise outcomes that are assessed. However, the 63 meta-analyses considered by Matt and Navarro did not permit clear conclusions about the influence that such factors might have on therapeutic effectiveness.

Discussion points

- 1. What are the strengths and weaknesses of meta-analyses as a way of discovering the effectiveness of therapy?
- 2. How impressed are you by the apparent effectiveness of most forms of therapy revealed by Matt and Navarro?

KEY STUDY EVALUATION

A major criticism of Matt and Navarro’s research concerns its lack of standardization. The specific conditions of individual cases are vital in determining the effectiveness of therapy. Linked to this is the timescale used. Regression or relapse would indicate failure of therapy, but this is impossible to discover from Matt and Navarro’s study. A much lengthier approach, capable of handling disparate sets of data, with a methodology using detailed case notes and follow-up research would be needed to address this problem.

Accuracy is difficult to determine in meta-analyses, and it would be of value to consider making more use of the case-study-centered approach. However, the inevitable cost and time implications militate against this more focused type of study.

Effective treatments for various mental disorders (based on information in Roth & Fonagy, 2005)

Disorder	Clearly effective	Limited support for efficacy
Major depressive disorder	Cognitive-behavior therapy Interpersonal psychotherapy	Short-term psychodynamic therapy
Social phobia	Exposure therapy Exposure therapy + cognitive therapy	
Generalized anxiety disorder	Cognitive-behavior therapy Applied relaxation	
Panic disorder (+with agoraphobia)	Exposure therapy Cognitive-behavior therapy Panic control therapy	
Posttraumatic stress disorder	Cognitive-behavior therapy Eye movement desensitization	Structured psychodynamic therapy
Specific phobia	Exposure therapy	
Schizophrenia	Family intervention programs	Cognitive therapy for delusions Psychoeducation + psychological intervention

Roth and Fonagy (2005) analyzed exhaustively the effectiveness of different forms of therapy. For each disorder, they identified those forms of treatment that have been shown to be clearly effective, and those that currently have limited support (see the box above). Family intervention programs were identified as clearly effective in the treatment of schizophrenia. These programs involve schizophrenics and their families devising

appropriate strategies and working together to achieve certain objectives. These programs have proved effective, but recent research suggests that the benefits are often greatly reduced over a 5-year period after treatment (Montero et al., 2006).

Roth and Fonagy erred on the side of caution, since there is reasonable evidence for the effectiveness of forms of therapy not emphasized by them. For example, psychodynamic therapy has proved effective in the treatment of schizophrenic patients having an initially high level of global functioning (Hauff et al., 2002). In addition, recent research indicates drug therapy is as effective as cognitive-behavior therapy in the treatment of generalized anxiety disorder (Mitte, 2005) and in the treatment of depression and several anxiety disorders (Otto et al., 2006).

WHY IS THERAPY EFFECTIVE?

There are two main reasons why any given therapy might be effective. First, there are **specific factors**, which are aspects of therapy specific or unique to that form of therapy. For example, Freud may have been correct in assuming that recovery will follow when patients gain insight into the nature of their problems. Second, there are **common factors**, which are aspects (e.g., therapist warmth; therapist empathy; therapeutic alliance between therapist and client) common to most forms of therapy.

Matt and Navarro (1997) considered 10 meta-analyses in which three types of group were compared:

1. Specific therapy groups, for whom any benefits may depend on specific effects or common effects.
2. Placebo control groups (involving general encouragement but no specific therapy), for whom any benefits are likely to depend on common effects.
3. Waiting-list control groups, for whom no benefits are expected.

The evidence indicated that 57% of placebo control patients did better than the average waiting-list control patient, indicating that common or placebo effects exist. However, 75% of the patients receiving specific therapy did better than the average placebo control patient, indicating that specific effects are much more powerful than common ones.

Stevens, Hynan, and Allen et al. (2000) reported similar findings in an analysis of 80 outcome studies in which specific therapy groups, common factor or placebo groups, and waiting-list control groups were compared. They also carried out a further analysis taking account of the severity of the mental disorder. With the less severe disorders, the impact of specific and common factors was comparable. With the more severe disorders, in contrast, *only* specific factors influenced the outcome. These findings make sense—a therapist *only* needs to be friendly and sympathetic to assist patients with minor disorders, but this is not enough to help patients with severe problems.

What characteristics should therapists possess (other than being friendly) to maximize their effectiveness? Lafferty, Beutler, and Crago (1989) found that effective therapists had greater empathy with or emotional understanding of their clients. In addition, effective therapists provided greater directiveness and support during therapy.

Najavits and Strupp (1994) considered the behavior of effective and ineffective therapists during treatment. They distinguished between positive behaviors (e.g., warmth, understanding, helping) and negative behaviors (e.g., ignoring, rejecting, and attacking). The effective therapists showed more positive behaviors and fewer negative behaviors than did the ineffective ones.



...warmth, acceptance and empathy on the part of the therapist.

Key Terms

Specific factors:

features unique to a given form of therapy that help the patient to recover; see **common factors**.

Common factors:

general factors found in most forms of therapy (e.g., therapist's personal qualities) that help the client to recover; see **specific factors**.

THE FUTURE?

If the truth be told, the picture I have painted so far is oversimplified. The notion that most therapists rigidly stick to a single therapeutic approach is increasingly incorrect. In reality, therapists typically practice **eclecticism**, using techniques drawn from various types of therapy rather than just one approach. As Kopta, Lueger, Saunders, and Howard (1999, p. 455) pointed out, “An increasingly popular view is that the long-term dominance of the major psychotherapies has ended and that integrationism and eclecticism is now the direction for technical advances in treatment.”

Several years ago, Prochaska and Norcross (1994) found that 38% of American therapists identified their approach as eclectic. This can be compared with 33% using only a psychodynamic approach, 5% using only cognitive therapy, and 5% using only behavior therapy. Unfortunately, relatively few studies have considered the effectiveness of eclectic approaches to treatment. A potential warning was sounded by Otto, Smits, and Reese (2006), who considered research in which anxiety disorders were treated by combined drug therapy and cognitive-behavior therapy. They concluded as follows: “In the anxiety disorders, there are some benefits [to combined therapy] in the short-term, but combined treatment may limit the maintenance of treatment gains offered by CBT [cognitive-behavior therapy] alone” (Otto et al., 2006, p. 72).

Key Term

Eclecticism:
the use of a range of different forms of treatment by a therapist rather than favoring a single therapeutic approach.

Chapter Summary

Drug therapy

- Drug therapy is used to treat mental disorders. Selective serotonin reuptake inhibitors (SSRIs) are commonly used to treat depression, the benzodiazepines are used for anxiety, and atypical antipsychotic drugs are used with schizophrenics. Relapse rates can be high because the drugs don't deal directly with the underlying problems.

Psychodynamic therapy

- According to Freud, successful therapy involves the client gaining insight into his/her repressed ideas and conflicts from childhood via free association and dream analysis. The psychodynamic approach triggered by Freud's psychoanalytic approach is moderately successful. However, it is unclear whether it works for the reasons proposed by Freud.

Behavior therapy

- Behavior therapists argue that classical and operant conditioning produce undesirable behavior and that conditioning can be used to produce desirable behavioral changes. Behavior therapy includes exposure therapy (which has proved very successful). It also includes aversion therapy and the token economy, both of which often fail to work in the long-term.

Cognitive and cognitive-behavior therapy

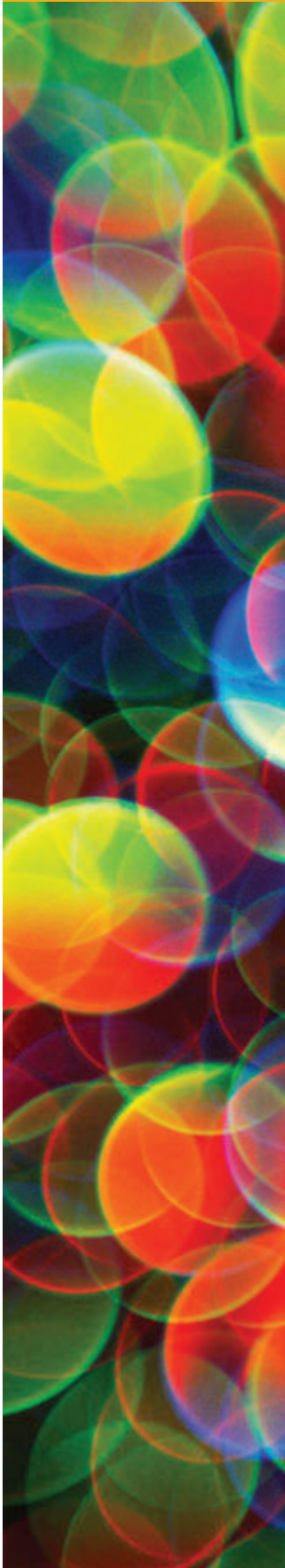
- According to the cognitive approach, patients suffering from mental disorders have distorted beliefs. In depressed patients, such thoughts center on the cognitive triad. In anxious patients, such thoughts center on exaggerating the threateningness of external and/or internal stimuli.
- Anxious patients' distorted beliefs are maintained because patients use safety-seeking behaviors that are mistakenly thought to prevent catastrophe.
- According to cognitive therapists, therapy should focus on eliminating patients' false beliefs. According to cognitive-behavior therapists, therapy should focus on changing behavior as well as eliminating false beliefs.
- Cognitive and cognitive-behavior therapy seem to be more effective in the treatment of anxiety and depression than more severe disorders such as schizophrenia.

Effectiveness of therapy

- Problems in assessing the effectiveness of therapy include selection of outcome measures, differences between efficacy and effectiveness studies, and the occurrence of nontherapist sources of assistance.
- Meta-analyses typically indicate that all major forms of therapy are moderately effective and differ only slightly in terms of effectiveness. However, there is increasing evidence that cognitive-behavior therapy is especially effective in treating depression and several anxiety disorders.
- Therapeutic effectiveness depends on specific and common factors. Both kinds of factors are important with less severe disorders, but specific factors are much more important than common ones with severe disorders.
- Effective therapists have high levels of empathy and provide more directiveness and support during therapy than less effective ones.
- Therapists increasingly use an eclectic approach, using techniques drawn from several different therapeutic approaches. Relatively little is known of the value of this approach.

Further Reading

- Butler, A.C., Chapman, J.E., Forman, E.M., & Beck, A.T. (2005). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review*, 26, 17–31. Convincing evidence that cognitive-behavior therapy is effective in the treatment of many disorders is presented in this review article.
- Comer, R.J. (2007). *Abnormal psychology* (6th ed.). New York: Worth. Most of the important issues are discussed in a thorough and accessible way in this up-to-date textbook.
- Durand, V.M., & Barlow, D.H. (2006). *Essentials of abnormal psychology* (4th ed.). New York: Thomson/Wadsworth. As the title implies, this textbook by leading clinical psychologists focuses on central issues within abnormal psychology.
- Jarvis, M. (2004). *Psychodynamic psychology: Classical theory and contemporary research*. London: Thomson. Matt Jarvis provides an entertaining and balanced view of the entire psychodynamic approach.
- Roth, A., & Fonagy, P. (with contributions from G. Parry, M. Target, & R. Woods) (2005). *What works for whom? A critical review of psychotherapy research* (2nd ed.). New York: Guilford Press. The authors of this valuable book provide comprehensive coverage of the factors determining the effectiveness of different forms of therapy.



- **Psychological inquiry** pp.541–559

- Experimental method*
- Quasi-experiments*
- Correlational studies*
- Naturalistic observation*
- Case studies*
- Interviews*
- Discourse analysis*

- **Design of investigations** pp.561–583

- Aims and hypotheses*
- Selecting participants*
- Good practice in experimentation*
- Experimental designs*
- Good practice in nonexperimental designs*
- Problems with experimental research*
- General issues in investigations*

- **Data analysis** pp.585–615

- Qualitative analysis of data*
- Interpretation of interviews, case studies, and observations*
- Content analysis*
- Quantitative analysis: Descriptive statistics*
- Data presentation*
- Statistical tests*
- Issues of experimental and ecological validity*
- Writing up a practical*

INTRODUCTION TO Research Methods

P psychology is basically (and increasingly) a scientific subject. It is thus extremely important for experiments and other studies in psychology to be carried out in as scientific a fashion as possible. Thus, an understanding of research methods is crucial for students of psychology.

ORGANIZATION OF CHAPTERS 23–25

Chapter 23 is concerned with some of the main research methods used by psychologists. You may be surprised to discover how many types of study there are. The “gold standard” involves use of the experimental method. However, there are also correlational studies, naturalistic observations, in-depth case studies, interview-based studies, and discourse analysis. Examples of all of these are provided and discussed.

Chapter 24 is concerned with a consideration of how to carry out experiments and other studies effectively. Relevant issues include deciding on the hypotheses to be tested, selecting participants, designing the experiment so as to avoid unwanted variables and errors, and deciding on the experimental design. Good experimental research involves avoiding various problems that can invalidate it. For example, the experimenter may unwittingly influence the participants’ behavior. This and other problems are discussed. There is also a section on issues relating to the design of various kinds of non-experimental studies including interviews, surveys, and case studies.

Chapter 25 focuses on data analysis. After we have collected data from an experimental or nonexperimental study, it is necessary to interpret those data and decide what conclusions are warranted. There are many ways in which data can be analyzed in order to provide a basis for interpreting them. There is a major distinction between *quantitative* analyses and *qualitative* analyses. Quantitative analyses use numerical data (e.g., reaction times; percentage correct responses), whereas qualitative analyses use non-numerical data (e.g., transcripts of interviews; information obtained from a case study).

An important goal of data analysis is to communicate findings to other people, and the use of charts and graphs for that purpose is discussed in Chapter 25. Other key issues are deciding *which* statistical test to use when quantitative analysis is called for, and then knowing *how* to perform the necessary calculations. Much experimental research is carried out under laboratory conditions, and an important issue is whether laboratory findings will generalize to everyday life (whether they possess ecological validity).

In sum, the main issues involved in carrying out psychological studies successfully are dealt with in this section. After reading all three chapters, you should be well equipped to carry out your own studies in psychology! I hope you find the process of carrying out and analyzing your own studies as fascinating as I have with my own research.

chapter 23

Contents

Experimental method	541
Quasi-experiments	546
Correlational studies	548
Naturalistic observation	549
Case studies	551
Interviews	553
Discourse analysis	555

Psychological inquiry

23

In common with other sciences, psychology is concerned with theories and with data. A **theory** provides a general explanation or account of certain findings or data. It also generates a number of **experimental hypotheses**, which are predictions or expectations about behavior based on the theory. For example, someone might propose a theory in which it is argued that some people are more hostile than others. This theory could be used to produce various hypotheses or predictions, such as the following: Hostile people will express anger more often than nonhostile ones; hostile people will react more strongly than nonhostile ones to frustrating situations; hostile people will be more sarcastic than nonhostile people.

Psychologists spend a lot of their time collecting data in the form of measures of behavior. Data are collected in order to test various hypotheses. Most people assume that this data collection involves proper or true experiments carried out under laboratory conditions, and it is true that literally millions of laboratory experiments have been carried out in psychology. However, psychologists make use of several methods of investigation, each of which has provided useful information about human behavior.

As you read through the various methods of investigation, it is natural to wonder which methods are the best and which are the worst. In some ways, it may be more useful to compare the methods used by psychologists to the clubs used by the golf professional. The driver is not a better or worse club than the putter, it is simply used for a different purpose. In similar fashion, each method of investigation used by psychologists is very useful for testing some hypotheses, but is of little or no use for testing other hypotheses. However, as we will see, the experimental method provides the best way of being able to make inferences about cause and effect.

EXPERIMENTAL METHOD

The method of investigation used most often by psychologists is the experimental method. In order to understand what is involved in the experimental method, we will consider a concrete example.

DEPENDENT AND INDEPENDENT VARIABLES

Suppose that a psychologist wants to test the experimental hypothesis that loud noise will have a disruptive effect on the performance of a task. As with most hypotheses, this one refers to a **dependent variable**, which is some aspect of behavior that is going to be measured. In this case, the dependent variable is some measure of task performance.

Most experimental hypotheses state that the dependent variable will be affected systematically by some specified factor, which is known as the **independent variable**. In the case we are considering, the independent variable is the intensity of noise. More generally, the independent variable is some aspect of the experimental situation that is manipulated by the experimenter.

We come now to the most important principle involved in the use of the experimental method: the independent variable of interest is manipulated, but all other variables are *controlled*. It is assumed that, with all other variables controlled, the one and only

Key Terms

Theory:

a general explanation of a set of findings; it is used to produce experimental hypotheses.

Experimental hypotheses:

the testable predictions generated by a theory; these usually specify independent and dependent variables.

Dependent variable:

some aspect of the participant's behavior that is measured in a study.

Independent variable:

some aspect of the experimental situation that is manipulated by the experimenter.

The experimental process can be summarized thus:

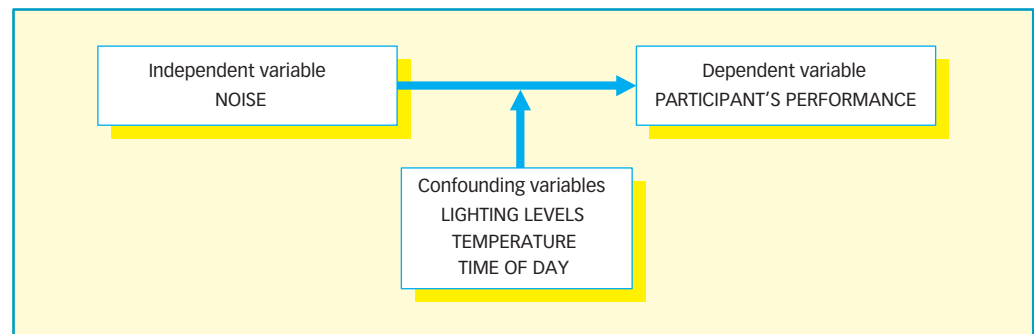
- Experimenter acts on independent variable
- Changes in independent variable lead to changes in dependent variable
- Changes in dependent variable measured by experimenter

variable that is being manipulated is the cause of any subsequent change in the dependent variable. In terms of our example, we might expose one group of participants to very intense noise, and a second group to mild noise. What would we need to do to ensure that any difference in the performance of the two groups was a result of the noise rather than any other factor? We would control all other aspects of the situation by, for example, always using the

same room for the experiment, keeping the temperature the same, and having the same lighting.

CONFOUNDING VARIABLES

Another way of expressing the essence of the experimental method is that it is of fundamental importance to avoid any **confounding variables**. These are variables that are manipulated or allowed to vary systematically along with the independent variable. The presence of any confounding variables has grave consequences, because it prevents us from being able to interpret our findings. For example, suppose that the participants exposed to intense noise performed the task in poor lighting conditions so that they could hardly see what they were doing, whereas those exposed to mild noise enjoyed good lighting conditions. If the former group performed much worse than the latter group, we would not know whether this was because of the intense noise, the poor lighting, or some combination of the two.



You might think that it would be easy to ensure that there were no confounding variables in an experiment. However, there are many well-known published experiments containing confounding variables. Consider, for example, a study by Jenkins and Dallenbach (1924). They gave a learning task to a group of participants in the morning, and then tested their memory for the material later in the day. The same learning task was given to a second group of participants in the evening, and their memory was tested the following morning after a night's sleep.

What did Jenkins and Dallenbach find? Memory performance was much higher for the second group than for the first. They argued that this was because there is less interference with memory when people are asleep than when they are awake. Can you see the flaw in this argument? The two groups learned the material at different times of day, and so time of day was a confounding variable. Hockey, Davies, and Gray (1972) discovered many years later that the time of day at which learning occurs is probably more important than whether or not the participants sleep between learning and the memory test.

Key Term

Confounding variables: variables that are mistakenly manipulated or allowed to vary along with the independent variable.

PARTICIPANTS AND SETTINGS

Proper use of the experimental method requires careful consideration of the ways in which the participants are allocated to the various conditions. A detailed account is given in Chapter 24, so we will focus here only on experiments in which there are different participants in each condition. Suppose that the participants exposed to intense noise were on average much less intelligent than those exposed to mild noise. We would then

be unable to tell whether poorer performance by the former participants was a result of the intense noise or their low intelligence. The main way of guarding against this possibility is by means of **randomization**, in which the participants are allocated at random to the two conditions.

Numerous studies are carried out using students as participants. This raises the issue of whether students are representative of society as a whole. For example, it is possible that students would be less distracted than other people by intense noise because they are used to studying over long periods of time in conditions that can be noisy, such as halls of residence.

The experimental method is used mainly in laboratory experiments. However, it is also used in **field experiments**, which are experiments carried out in natural settings such as in the street, in a school, or at work. Some of the advantages of the experimental method are common to both laboratory and field experiments, whereas other advantages and limitations are specific to one type of experiment. We will consider the common advantages next, with more specific advantages and limitations being discussed after that.

In many studies, use is made of pre-existing groups of people. For example, we might compare the performance of males and females, or that of young and middle-aged individuals. Do such studies qualify as genuine experiments? The answer is “No.” Use of the experimental method requires that the independent variable is *manipulated* by the experimenter, but clearly the experimenter cannot decide whether a given person is going to be male or female for the purposes of the study!



Ideally, psychological experiments should select a random sample of the population, although true randomness can be hard to achieve.

COMMON ADVANTAGES

Causal relationships

What is generally regarded as the greatest advantage of the experimental method is that it allows us to establish cause and effect relationships. In the terms we have been using, the independent variable in an experiment is often regarded as a cause, and the dependent variable is the effect. Philosophers of science have argued about whether or not causality can be established by experimentation. However, the general opinion is that causality can only be inferred. If y (e.g. poor performance) follows x (e.g. intense noise), then it is reasonable to infer that x caused y .

We can see why findings from studies based on the experimental method do not necessarily establish causality from the following imaginary example. An experiment on malaria is carried out in a hot country. Half of the participants sleep in bedrooms with the windows open, and the other half sleep in bedrooms with the windows closed. Those sleeping in bedrooms with the windows open are found to be more likely to catch malaria. It would obviously be wrong to argue that having the window open caused malaria. Having the window open or closed is relevant to catching the disease, but it tells us nothing directly about the major causal factor in malaria (infected mosquitoes).

Replication

The other major advantage of the experimental method concerns what is known as **replication**. If an experiment has been conducted in a carefully controlled way, it should be possible for other researchers to repeat or replicate the findings obtained from that experiment. There have been numerous failures to replicate using the experimental method, but the essential point is that the chances of replication are greater when the experimental method is used than when it is not.

LABORATORY VS. FIELD EXPERIMENTS

Laboratory and field experiments both involve use of the experimental method, but they differ in that field experiments are carried out in more natural settings. As an example of a field experiment, let us consider a study by Shotland and Straw (1976; see p. 433).

Key Terms

Randomization:

the allocation of participants to conditions on a random basis.

Field experiment:

a study in which the experimental method is used in a naturalistic situation.

Replication:

the ability to repeat the findings obtained from an experiment.



When bystanders saw a staged fight in an experiment by Shotland and Straw (1976) they were more likely to help when they thought two strangers were involved than if they thought the couple were married.

They arranged for a man and a woman to stage an argument and a fight fairly close to a number of bystanders. In one condition, the woman screamed, “I don’t know you.” In a second condition, she screamed, “I don’t know why I ever married you!” When the bystanders thought the fight involved strangers, 65% of them intervened, against only 19% when they thought it involved a married couple. Thus, people were less likely to lend a helping hand when it was a “lovers’ quarrel” than when it was not. The bystanders were convinced that the fight was genuine, as was shown by the fact that 30% of the women were so alarmed that they shut the doors of their rooms, turned off the lights, and locked their doors.

The greatest advantage of laboratory experiments over field experiments is that it is generally easier to eliminate confounding variables in the laboratory than in the field. The experimenter is unlikely to be able to control every aspect of a natural situation.

Another clear advantage of laboratory experiments over field experiments is that it is much easier to obtain large amounts of very detailed information from participants in the laboratory. For example, it is hard to see how information about participants’ physiological activity or speed of performing a range of complex cognitive tasks could be obtained in a field experiment carried out in a natural setting. There are two main reasons why field experiments are limited in this way. First, it is not generally possible to introduce bulky equipment into a natural setting. Second, the participants in a field experiment are likely to realize they are taking part in an experiment if attempts are made to obtain a lot of information from them.

One of the advantages of field experiments over laboratory experiments is that the behavior of the participants is often more *typical* of their normal behavior. However, the greatest advantage of field experiments over laboratory experiments is that they are less artificial. The artificiality of laboratory experimentation was emphasized by Heather (1976, pp. 31–33):

Psychologists have attempted to squeeze the study of human life into a laboratory situation where it becomes unrecognizably different from its naturally occurring form . . . Experiments in psychology . . . are social situations involving strangers, and it might be suggested that the main kind of knowledge gleaned from years of experimentation with human subjects is information about how strangers interact in the highly artificial and unusual setting of the psychological experiment.

The effects of being observed

An important reason why laboratory experiments are more artificial than field experiments is because the participants in laboratory experiments are aware that their behavior is being observed. As Silverman (1977) pointed out, “Virtually the only condition in which a subject [participant] in a psychological study will not behave as a subject [participant] is if he does not know he is in one.” One consequence of being observed is that the participants try to work out the experimenter’s hypothesis, and then act accordingly. In this connection, Orne (1962) emphasized the importance of **demand characteristics**, which are “the totality of cues which convey an experimental hypothesis to the subjects.” Orne found that the participants in one of his studies were willing to spend several hours adding numbers on random number sheets and then tearing up each completed sheet. Presumably the participants interpreted the experiment as a test of endurance, and this motivated them to keep going.

Another consequence of the participants in laboratory experiments knowing they are being observed is **evaluation apprehension**. Rosenberg (1965) defined this as “an active anxiety-toned concern that he [the participant] win a positive evaluation from the experimenter or at least that he provide no grounds for a negative one.”

Key Terms

Demand characteristics: cues used by participants to try to guess the nature of the study or to work out what the experiment is about.

Evaluation apprehension: anxiety-toned concern felt by participants to perform well and please the experimenter.

Sigall, Aronson, and Van Hoose (1970) contrasted the effects of demand characteristics and evaluation apprehension on the task of copying telephone numbers. The experimenter told participants doing the task for the second time that he expected them to perform it at a rate that was actually slower than their previous performance. Adherence to the demand characteristics would have led to slow performance, whereas evaluation apprehension and the need to be capable would have produced fast times. The participants actually performed more quickly than they had done before, indicating the greater importance of evaluation apprehension.

This conclusion was strengthened by the findings from a second condition, in which the experimenter not only said that he expected the participants to perform at a slower rate, but also told them that those who rush are probably obsessive-compulsive. The participants in this condition performed the task slowly, because they wanted to be evaluated positively.

Another way in which laboratory experiments tend to be more artificial than field experiments was identified by Wachtel (1973). He used the term **implacable experimenter** to describe the typical laboratory situation, in which the experimenter's behavior (e.g., instructions) affects the participant's behavior but the participant's behavior does not influence the experimenter's behavior. There are two problems with experiments using an implacable or unyielding experimenter. First, because the situation (including the experimenter) is allowed to influence the participant but the participant isn't allowed to affect the situation, it is likely that the effects of situations on our behavior are overestimated. Second, because much of the richness of the dynamic interactions between individual and situation has been omitted, there is a real danger that seriously oversimplified accounts of human behavior will emerge.

Artificiality

How much does it matter that laboratory experiments are artificial? As Coolican (1998) pointed out, "In scientific investigation, it is often *necessary* to create artificial circumstances in order to *isolate* a hypothesized effect." If we are interested in studying basic cognitive processes such as those involved in perception or attention, then the artificiality of the laboratory is unlikely to affect the results. On the other hand, if we are interested in studying social behavior, then the issue of artificiality does matter. For example, Zegoib, Arnold, and Forehand (1975) found that mothers behaved in a warmer and more patient way with their children when they knew they were being observed than when they did not.

Carlsmith, Ellsworth, and Aronson (1976) drew a distinction between **mundane realism** and **experimental realism**. Mundane realism refers to experiments in which the situation is set up to resemble situations often found in everyday life. In contrast, experimental realism refers to experiments in which the situation may be rather artificial, but is sufficiently interesting to produce full involvement from the participants. Milgram's (1974) research on obedience to authority is a good example of experimental realism (see Chapter 19). The key point is that experimental realism may be more important than mundane realism in producing findings that generalize to real-life situations.



Participants in psychological experiments usually try to perform the task set by the experimenter as well as they can, in order to gain his or her approval.

Laboratory experiments

Advantages

- Establishes cause-and-effect relationships
- Allows for replication
- Good control of confounding variables

Limitations

- Artificial
- Participants know they are being observed (demand characteristics and evaluation apprehension)
- Low in external validity
- Ethical concerns, such as the right to withdraw

Key Terms

Implacable experimenter:

the typical laboratory situation in which the experimenter's behavior is uninfluenced by the participant's behavior.

Mundane realism:

the use of an artificial situation that closely resembles a natural situation.

Experimental realism:

the use of an artificial situation in which the participants become fully involved.

Field experiments

Advantages

- Establishes cause and effect relationships
- Allows for replication
- Behavior of participants more typical than in a laboratory experiment: high external validity

Limitations

- Ethical issues, such as a lack of voluntary informed consent
- Low in internal validity: poor control

Ethical issues

What we will do here is discuss a few ethical issues that are of special relevance to laboratory or field experiments. So far as laboratory experiments are concerned, there is a danger that the participants will be willing to behave in a laboratory in ways they would not behave elsewhere. For example, Milgram (1974) found in his work on obedience (see Chapter 19) that 65% of his participants were prepared to give very intense electric shocks to someone else when the experiment took place in a laboratory at Yale University. In contrast, the figure was only 48% when the same study was carried out in a run-down office building.

Thus, participants are often willing to do what they would not normally do in the setting of a prestigious laboratory.

Another ethical issue that applies especially to laboratory experiments concerns the participant's right to withdraw from the experiment at any time. It is general practice to inform participants of this right at the start of the experiment. However, participants may feel reluctant to exercise this right if they think it will cause serious disruption to the experimenter's research.

So far as field experiments are concerned, the main ethical issue relates to the principle of voluntary informed consent, which is regarded as central to ethical human research. By their very nature, most field experiments do not lend themselves to obtaining informed consent from the participants. For example, the study by Shotland and Straw (1976) would have been rendered almost meaningless if the participants had been asked beforehand to give their consent to witnessing a staged quarrel! The participants in that study could reasonably have complained about being exposed to a violent quarrel.

Another ethical issue with field experiments is that it is not possible in most field experiments to tell the participants that they have the right to withdraw at any time without offering a reason.

Summary

The respective strengths and weaknesses of laboratory experiments and field experiments can be summed up with reference to two different kinds of validity: internal validity and external validity. Internal validity refers to the validity of an experiment within the confines of the context in which it is carried out, whereas external validity refers to the validity of an experiment outside the research situation itself. Laboratory experiments tend to be high in internal validity but low in external validity, whereas field experiments are high in external validity but low in internal validity.

QUASI-EXPERIMENTS

"True" experiments based on the experimental method provide the best way of being able to draw causal inferences with confidence. However, it is often the case that there are practical or ethical reasons why it is simply not possible to carry out a true experiment. In such circumstances, investigators often carry out what is known as a **quasi-experiment**. Quasi-experimental designs "resemble experiments but are weak on some of the characteristics" (Raulin & Graziano, 1994). There are two main ways in which quasi-experiments tend to fall short of being true experiments. First, the manipulation of the independent variable is often not under the control of the experimenter. Second, it is usually not possible to allocate the participants randomly to groups.

There are numerous hypotheses in psychology that can only be studied by means of quasi-experiments rather than true experiments. For example, suppose that we are interested in studying the effects of divorce on young children. We could do this by comparing children whose parents had divorced with those whose parents were still married. There would, of course, be no possibility of allocating children at random to the

Key Term

Quasi-experiment: a type of experiment resembling a "true" experiment, but with some aspects of the experimental method omitted.

divorced or non-divorced parent groups! Studies in which pre-existing groups are compared often qualify as quasi-experiments. Examples of such quasi-experiments would be comparing the learning performance of males and females, or comparing the social behavior of introverted and extraverted individuals.

NATURAL EXPERIMENTS

The **natural experiment** is a type of quasi-experiment in which a researcher makes use of some naturally occurring event for research purposes. An example of a natural experiment is a study by Williams (1986) on the effects of television on aggressive behavior in Canadian children aged between 6 and 11 years. Three communities were compared: one in which television had just been introduced, one in which there was only one television channel, and one in which there were several channels. The children in the first community showed a significant increase in verbal and physical aggression during the first 2 years after television was introduced, whereas those in the other two communities did not. This was not a true experiment, because the children were not allocated randomly to the three conditions or communities.

Adams and Adams (1984) carried out a natural experiment following the eruption of the Mount St. Helens volcano in 1980. As the volcanic eruption had been predicted, they were able to assess the inhabitants of the small town of Othello before and after it happened. There was a 50% increase in mental health appointments, a 198% increase in stress-aggravated illness, and a 235% increase in diagnoses of mental illness.

Advantages and limitations

What are the advantages of natural experiments? The main one is that the participants in natural experiments are often not aware that they are taking part in an experiment, even though they are likely to know that their behavior is being observed. Another advantage of natural experiments is that they allow us to study the effects on behavior of independent variables that it would be unethical for the experimenter to manipulate. For example, Adams and Adams (1984) were interested in observing the effects of a major stressor on physical and mental illness. No ethical committee would have allowed them to expose their participants deliberately to stressors that might cause mental illness, but they were able to take advantage of a natural disaster to conduct a natural experiment.

What are the limitations of natural experiments? The greatest limitation occurs because the participants have not been assigned at random to conditions. As a result, observed differences in behavior between groups may be a result of differences in the types of participants in the groups rather than the effects of the independent variable. Consider, for example, the study by Williams (1986) on television and aggression. The children in the community that had just been exposed to television might have been naturally more aggressive than the children in the other two communities. However, the children in the three communities did not differ in their level of aggression at the start of the study.

It is usually possible to check whether the participants in the various conditions are comparable. For example, they can be compared with respect to variables such as age, sex, socioeconomic status, and so on. If the groups do differ significantly in some respects irrelevant to the independent variable, then this greatly complicates the task of interpreting the findings of a natural experiment.

The other major limitation of natural experiments involves the independent variable. In some natural experiments, it is hard to know exactly what aspects of the independent variable have caused any effects on behavior. For example, there is no doubt that the eruption of Mount St. Helens was a major stressor. It caused stress in part because of the



Adams and Adams (1984) designed a natural experiment around the eruption of the Mount St. Helens volcano in which they assessed the effects of stress on the population of a small town threatened by the eruption.

Key Term

Natural experiment: a type of quasi-experiment in which use is made of some naturally occurring event.

Quasi-experiments*Advantages*

- Participants behave naturally
- Investigates the effects of independent variables that it would be unethical to manipulate

Limitations

- Participants not allocated at random to conditions
- Difficult to identify what aspects of the independent variable have caused the effects on behavior

possibility that it might erupt again and produce more physical devastation. However, social factors were also probably involved. If people in Othello observed that one of their neighbors was highly anxious because of the eruption, this may have heightened their level of anxiety.

Ethical issues

It can be argued that there are fewer ethical issues with natural experiments than with many other kinds of research. The reason is that the experimenter is not responsible for the fact that the participants have been exposed to the independent variable. However, natural

experiments can raise various ethical issues. First, there can be the issue of informed voluntary consent, in view of the fact that the participants are often not aware that they are taking part in an experiment. Second, experimenters carrying out natural experiments need to be sensitive to the situation in which the participants find themselves. People who have been exposed to a natural disaster such as a volcanic eruption may resent it if experimenters start asking them detailed questions about their mental health or psychological wellbeing.

CORRELATIONAL STUDIES

Suppose that we were interested in the hypothesis that watching violence on television leads to aggressive behavior. One way of testing this hypothesis would be to obtain two kinds of information from a large number of people: (1) the amount of violent television they watched; and (2) the extent to which they behaved aggressively in various situations. If the hypothesis is correct, then we would expect that those who have seen the most violence on television would tend to be the most aggressive. In technical terms, this study would be looking for a **correlation**, or association, between watching violent programs and being aggressive. Thus, the closer the link between them, the greater would be the correlation or association.

One of the best-known uses of the correlational approach is in the study of the role of nature and nurture in intelligence. Similarity in intelligence within pairs is worked out by means of a correlation. Identical twins are more alike genetically than fraternal twins. As a result, their levels of intelligence should be more similar than those of fraternal twins if heredity plays an important role in determining intelligence. As predicted, the correlation indicating the degree of similarity in intelligence is nearly always higher for identical twins than for fraternal twins. However, it has proved hard to provide a detailed interpretation of the findings.

ADVANTAGES AND LIMITATIONS

Correlational designs are generally regarded as inferior to experimental designs, because it is hard (or impossible) to establish cause and effect. In our example, the existence of an association between the amount of television violence watched and aggressive behavior would certainly be consistent with the hypothesis that watching violent programs can cause aggressive behavior. However, there are other possible interpretations of the data. The causality may actually operate in the opposite direction. In other words, aggressive individuals may choose to watch more violent programs than those who are less aggressive. There may be a third variable that accounts for the association between the variables of interest, that is, watching violent programs and aggressive behavior. For example, people in disadvantaged families may watch more television programs of all kinds than those in nondisadvantaged families, and their deprived circumstances may also cause them to behave aggressively. If that were the case, then the number of violent television programs watched might have no direct effect at all on aggressive behavior.

In spite of the interpretive problems posed by the findings of correlational studies, there are several reasons why psychologists continue to use this method.

Key Term

Correlation:
an association that is found
between two variables.

First, many hypotheses cannot be examined directly by means of experimental designs. For example, the hypothesis that smoking causes a number of physical diseases cannot be tested by forcing some people to smoke and forcing others not to smoke! All that can be done is to examine correlations or associations between the number of cigarettes smoked and the probability of suffering from various diseases.

Second, it is often possible to obtain large amounts of data on a number of variables in a correlational study much more rapidly and efficiently than would be possible using experimental designs. Use of a questionnaire, for example, would permit a researcher to investigate the associations between aggressive behavior and a wide range of activities (such as watching violent movies in the cinema; reading violent books; being frustrated at work or at home).

Third, interpretive problems are much reduced if there is no association between two variables. For example, if it were found that there was no association at all between the amount of violent television watched and aggressive behavior, this would provide fairly strong evidence that aggressive behavior is not caused by watching violent programs on television.

Fourth, the interpretive problems with associations or correlations between two variables are often not as great as in the example of violent programs and aggression. Suppose, for example, we discover a correlation between age and happiness, in which older people are generally less happy than younger people. Although it would not be possible to offer a definitive interpretation of this finding, we could be entirely confident that unhappiness does not cause old age!

Correlational studies

Advantages

- Allows study of hypotheses that cannot be examined directly
- More data on more variables can be collected more quickly than in an experimental set-up
- Problems of interpretation are reduced when no association is found
- Even when strong correlations are found it may be obvious that no causal relationship exists

Limitations

- Interpretation of results is difficult
- Cause and effect cannot be established
- Direction of causality is uncertain
- Variables other than the one of interest may be operating

ETHICAL ISSUES

Correlational analyses are used very widely. As a result, it is not possible to identify any particular ethical issues that apply to most studies in which such analyses are carried out. However, correlational analyses are often used in socially sensitive research, which raises political and/or social issues. For example, consider the correlational evidence suggesting that individual differences in intelligence depend in part on genetic factors. Some people have argued mistakenly that this implies that race differences in intelligence also depend on genetic factors. The key ethical issue here (and in many other correlational studies) is for the researcher to be fully aware of the social sensitivity of the findings that he or she has obtained.

Another ethical issue is raised by the real possibility that the public at large will misinterpret the findings from correlational studies. For example, the finding that there was a correlation between the amount of television violence watched by children and their level of aggression led many influential people to argue that television violence was having a damaging effect. In other words, they mistakenly supposed that correlational evidence can demonstrate a causal relation. Television companies may have suffered from such overinterpretation of findings.

NATURALISTIC OBSERVATION

Naturalistic observation involves methods designed to examine behavior without the experimenter interfering with it in any way. This approach was originally developed by the ethologists Lorenz and Tinbergen. They studied animals in their natural habitat rather than in the laboratory, and discovered much about their behavior. An example of the use of naturalistic observation in human



The issue of whether or not there is a correlation between violence on television and aggressive behavior is frequently debated in the media.

research is the work of Brown, Fraser, and Bellugi (1964). They studied the language development of three children (Adam, Eve, and Sarah) by visiting them at home about 35 times a year.

One of the key requirements of the method of naturalistic observation is to avoid *intrusion*. Dane (1994, p. 1149) defined this as “anything that lessens the participants’ perception of an event as natural.” There are various ways in which intrusion can occur. For example, there will be intrusion if observations are made in an environment that the participants regard as a research setting. There will also be intrusion if the participants are aware that they are being observed. In many studies, the experimenter is in the same room as the participants, and so they are almost certain to realize they are being observed. When this is the case, the experimenter may try to become a familiar and predictable part of the situation before any observations are recorded.

The participants in naturalistic observation often display a wide range of verbal and nonverbal behavior. How can observers avoid being overloaded in their attempts to record this behavior? One approach is to focus only on actions or events that are of particular interest to the researcher; this is known as event sampling. Another approach is known as time sampling, in which observations are only made during specified time periods (e.g., the first 10 minutes of each hour). A third approach is point sampling, in which one individual is observed in order to categorize their current behavior, after which a second individual is observed.

In considering the data obtained from naturalistic observation, it is important to distinguish between recording and interpretation or coding. For example, an observer may record that the participant has moved forward, and interpret that movement as an aggressive action. In practice, however, observers typically only focus on interpreting or coding the participants’ behavior. For example, Bales (1950) developed the interaction process analysis, which allows observers to record inferred meanings for the forms of behavior shown by members of a group (e.g., “offers suggestion”).

There have been various attempts to develop ways of categorizing people’s behavior in naturalistic observation without interpreting it. For example, McGrew (1972) devised a detailed and comprehensive recording system to place the social interactions of children at nursery school into 110 categories.

ADVANTAGES AND LIMITATIONS

What are the advantages of naturalistic observation? First, if the participants are unaware that they are being observed, then it provides a way of observing people behaving naturally. When this happens, there are no problems from demand characteristics, evaluation apprehension, the implacable experimenter, and so on. Second, many studies based on naturalistic observation provide richer and fuller information than typical laboratory experiments. For example, participants’ behavior may be observed in a range of different social contexts rather than on their own in the laboratory. Third, it is sometimes possible to use naturalistic observation when other methods cannot be used. For example, the participants may be unwilling to be interviewed or to complete a questionnaire. In the case of participants being observed at work, it may be impossible to obtain permission to disrupt their work in order to carry out an experiment.

What are the limitations of naturalistic observation? These are some of the major ones:

- The experimenter has essentially no control over the situation; this can make it very hard or impossible to decide what caused the participants to behave as they did.
- The participants are often aware that they are being observed, with the result that their behavior is not natural.
- There can be problems of reliability with the observational measures taken, because of bias on the part of the observer or because the categories into which behavior is coded are imprecise. Attempts to produce good reliability often involve the use of very precise but narrow categories, leading to much of the participants’ behavior simply being ignored. Reliability can be assessed by correlating the observational records of two

different observers. This produces a measure of inter-rater reliability.

- The fact that observations are typically interpreted or coded prior to analysis can cause problems with the validity of measurement. For example, it may be assumed invalidly that all instances of one child striking another child represent aggressive acts, when in fact many of them are only playful gestures. Thus, great care needs to be taken in **operationalization**, which is a procedure in which a variable (e.g., aggressive act) is defined by the operations taken to measure it.
- There are often problems of replication with studies of naturalistic observation. For example, the observed behavior of children in a school may depend in part on the fact that most of the teachers are very lenient and fail to impose discipline. The findings might be very different at another school in which the teachers are strict.

Naturalistic observation

Advantages

- People tend to behave naturally
- Information that is gathered is rich and full
- Can be used where other methods are not possible

Limitations

- Experimenter has no control over the situation
- Participants can be aware of being watched and this can affect behavior
- Problems of reliability as a result of bias or imprecise categorization of behavior
- Problems of validity resulting from observers' or coders' assumptions
- Replication is not usually possible

ETHICAL ISSUES

Naturalistic observation poses ethical problems if the participants do not realize that their behavior is being observed. In those circumstances, they obviously cannot give their voluntary informed consent to be involved in the study. There can also be problems about confidentiality. Suppose, for example, that naturalistic observation takes place in a particular school, and the published results indicate that many of the children are badly behaved. Even if the name of the school is not mentioned in the report, many people reading it will probably be able to identify the school because they know that the researchers made detailed observations there.

CASE STUDIES

The great majority of studies in psychology have involved the use of experimental or correlational methods on groups of participants. These approaches permit the use of statistical techniques providing information about the extent to which the results obtained from a given sample can be generalized to some larger population.

There are often good reasons why it is not feasible to use numerous participants in a study. For example, a brain-damaged patient may have a very unusual pattern of impaired performance, and there may not be other patients having the same pattern. Another example might be a therapist who has a patient with a rare mental disorder, but there is no possibility of him or her collecting data from other patients with the same disorder. In such circumstances, it can be very useful to carry out a **case study**, in which one individual is investigated thoroughly and over a period of time.

Some researchers have argued that the study of individual cases can be more fruitful than the study of groups of participants. One of the most convincing statements of that argument was put forward by Gordon Allport (1962):

Why should we not start with individual behavior as a source of hunches . . . and then seek our generalizations but finally come back to the individual not for the mechanical application of laws but for a fuller and more accurate assessment than we are now able to give? . . . We stop with our wobbly laws of generality and seldom confront them with the concrete person.

Some of those who have favored single-case studies have been of an anti-scientific persuasion. However, a prominent experimentalist who advocated the use of single-case studies was the behaviorist B.F. Skinner. In a discussion of research on operant conditioning, Skinner (1966) argued, “instead of studying a thousand rats for one hour each, or a hundred rats for ten hours each, the investigator is likely to study one rat for a thousand hours.”

Key Terms

Operationalization:

a procedure in which variables of interest are defined by the operations taken to measure them.

Case study:

detailed investigation of a single individual.



Can psychologists learn as much from the detailed study of the behavior of a single rat as they can from a more superficial study of a large number of rats?

There are several types of case studies, and they are carried out for various reasons, some of which will be considered here. One reason is to test a current theory. For example, Atkinson and Shiffrin (1968) argued that information only enters the long-term memory store via the short-term memory store (see Chapter 8). As a result, a brain-damaged individual with impaired short-term memory should also have impaired long-term memory. Evidence that seemed to be inconsistent with this theory was reported in a case study on KF, who was involved in a motorcycle accident (Shallice & Warrington, 1970). He had very poor short-term memory for words and digits, but his long-term learning and recall were unaffected.

Case studies can also be used to refine theories. Baddeley and Hitch (1974) argued that people possess a phonological loop that is used in the rehearsal of verbal information (see Chapter 8). It used to be assumed that rehearsal within the phonological loop requires use of the speech muscles. However, Baddeley and Wilson (1985) carried out a case study on a student, GB. He suffered from anarthria, which meant that he could not use his speech muscles and was unable to speak. In spite of this disorder, GB was able to make use of the articulatory loop.

Key Study

Freud: Developing new theories

Case studies are also of value in the development of new theories. For example, Sigmund Freud carried out a case study on Dr. Schreber, a lawyer who suffered from paranoia (a mental disorder involving delusions). Freud was puzzled by the fact that Dr. Schreber and other paranoid patients had a number of apparently unrelated delusions. These included a jealous feeling that their spouses or lovers had been unfaithful to them, the belief that others were plotting against them, the belief that several members of the opposite sex were in love with them, and delusions of grandeur.

Freud's discussions with Dr. Schreber led him to an analysis of paranoia. According to Freud, homosexual desires underlie paranoia. However, because homosexuality was viewed with disfavor by society at that time, Freud felt that these desires remain unconscious and become distorted in various ways. One such distortion is for a male paranoiac to think his wife or lover loves another man, which produces jealousy. Delusions of grandeur were accounted for by assuming that the male paranoid individual's thought "I love a man," turns into "I love no one," and then into "I love no one but myself." This analysis led Freud to develop a new (but implausible) theory of paranoia.

Discussion points

1. Can you spot any problems with the case study on Dr. Schreber?
2. Is there any way in which we could test Freud's proposed interpretation of the case study?



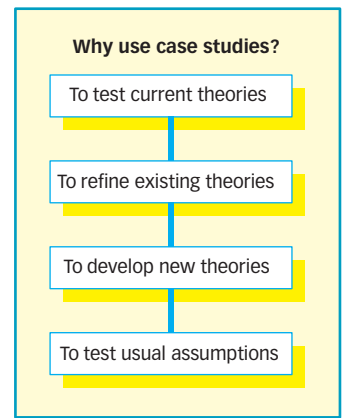
Actress Joanne Woodward portrayed the different "personalities" of Chris Sizemore in the film *The Three Faces of Eve*.

Some case studies are based on very unusual individuals. For example, there was Chris Sizemore, who was the central character in the film *The Three Faces of Eve*. Some of the time she was Eve White, a well-behaved and inhibited woman. At other times, she was Eve Black, who was promiscuous and impulsive. At still other times, she was Jane, who was more stable than either of the other two personalities. The existence of individuals with multiple personalities raises issues about our usual assumption that everyone has one personality and one self.

ADVANTAGES AND LIMITATIONS

What are the advantages of case studies? First, as we have seen, a single-case study can provide good evidence that a particular theory is in error. Of course, it is then desirable to find and test other individuals to check the findings from the first case study. Second, a case study can help to refine our theoretical understanding. Third, case studies can provide rich information that is used by the researcher or therapist to develop new theoretical ideas. An example of this is the case of Dr. Schreber, which was discussed in the Key Study earlier. Fourth, case studies can provide information about exceptional types of behavior or performance that had been thought to be impossible.

What are the limitations of case studies? The greatest limitation is their typically low reliability. The findings that are obtained from one unusual or exceptional individual are unlikely to be repeated in detail from another individual. Thus, it is often very hard to generalize from a single-case study. Second, many case studies involve the use of lengthy, fairly unstructured, interviews. Such case studies share the limitations that are identified for interviews in the next section. Third, researchers generally only report some of the data they obtained from their interviews with the participant. They may be unduly selective in terms of what they choose to report or to omit.



ETHICAL ISSUES

Case studies with clinical patients can pose important issues about confidentiality. A therapist such as Sigmund Freud may want to publish details of his case studies because they seem to support his theoretical position. However, the patient may be very unwilling for personal information about him or her to be published. Case studies with brain-damaged patients can pose ethical issues about voluntary informed consent. For example, patients with severe language impairments may find it hard to understand what will be involved in a case study, and so they cannot give proper informed consent.

INTERVIEWS

As Coolican (1994) pointed out, there are various kinds of interview, which vary enormously in terms of the amount of structure they contain. In what follows, we make use of his categorization of different types of interview.

NONDIRECTIVE INTERVIEWS

Nondirective interviews possess the least structure, with the person being interviewed (the interviewee) being free to discuss almost anything he or she wants. The role of the interviewer in nondirective interviews is to guide the discussion and to encourage the interviewee to be more forthcoming. This type of interview is used very often in psychotherapy, but has little relevance to research.

INFORMAL INTERVIEWS

Informal interviews resemble nondirective interviews, in that the interviewer listens patiently and focuses mainly on encouraging the interviewee to discuss issues in more depth or detail. However, informal interviews differ in that there are certain general topics that the interviewer wishes to explore. One of the best-known examples involving informal interviews was a large-scale study of workers at the Hawthorne works of Western Electric. The aim of this study was to explore industrial relations via a series of interviews. What emerged from informal interviews was that the relatively minor issues initially raised by the workers generally reflected deeper and more serious worries (Roethlisberger & Dickson, 1939; see p. 579).

GUIDED INTERVIEWS

Informal but guided interviews possess a little more structure than informal interviews. The interviewer identifies beforehand the issues to be addressed, but how and when to



There are various types of interviews used for psychological experimentation, from nondirective interviews to fully structured designs that have a standard set of questions with restricted-choice answers.

raise those issues is decided during the course of the interview. Structured but open-ended interviews use a formal procedure in which all interviewees are asked precisely the same questions in the same order. Such a procedure prevents the interviewee from side-tracking the interview and taking control of it away from the interviewer. The interviews are open-ended, in the sense that the questions that are asked allow plenty of scope for various kinds of answers (e.g. “How do you see your career developing?”).

CLINICAL INTERVIEWS

The clinical interview or clinical method resembles the structured but open-ended interview. In essence, all of the interviewees or participants are asked the same questions, but the choice of follow-up questions depends on the answers that are given. Piaget made much use of the clinical method in his research on cognitive development in

children (see Chapter 14). Piaget understood that children might perform poorly on a task because they did not understand fully what the experimenter wanted them to do. One way of trying to avoid this problem was by giving the experimenter the flexibility to ask questions in various ways. In spite of this, critics of Piaget have argued that the children he studied often failed to solve problems because of the complex language used by the experimenter.

FULLY STRUCTURED INTERVIEWS

Finally, there is the fully structured interview. In this type of interview, a standard set of questions is asked in the same fixed order to all of the interviewees, and they are only allowed to choose their answers from a restricted set of possibilities (e.g. “Yes”; “No”; “Don’t know”). As Coolican (1994, pp. 121–122) pointed out, “this approach is hardly an interview worth the name at all. It is a face-to-face data-gathering technique, but could be conducted by telephone or by post.”

ADVANTAGES AND LIMITATIONS

What are the advantages of the interview method? As might be expected, the precise advantages depend on the type of interview. Relatively unstructured interviews have the advantage that they are responsive to the personality, interests, and motivations of the interviewee. In principle, they can perhaps reveal more about the interviewee than is likely to be the case with more structured interviews. One of the advantages of fairly structured interviews is that it is easy to compare the responses of different interviewees, all of whom have been asked the same questions. Another advantage is good reliability, in that two different interviewers are likely to obtain similar responses from an interviewee when they ask exactly the same questions in the same order. A further advantage is that there is a reasonable probability of being able to replicate or repeat the findings from a study using structured interviews. Finally, structured interviews have the advantage that it is usually fairly easy to analyze the data obtained from them.

What are the limitations of the interview method? So far as unstructured interviews are concerned, there is the problem that the kinds of information obtained from different interviewees vary in an unsystematic way. As a result, the data from unstructured interviews tend to be hard to analyze. A further limitation with unstructured interviews is that what the interviewee says is determined in a complex way by the interaction between him or her and the interviewer. In other words, the personality and other characteristics of the interviewer typically influence the course of the interview, and make it hard to work out which of the interviewee’s contributions are and are not affected by

the interviewer. Finally, the fact that the information obtained from interviewees in unstructured interviews is influenced by the interviewer means that the data obtained can be viewed as unreliable.

One of the main limitations with structured interviews is that what the interviewee says may be somewhat constrained and artificial because of the high level of structure built into the interview. Another limitation is that there is little or none of the flexibility associated with unstructured interviews.

Finally, we need to consider three limitations that are common to all types of interview. First, there is the issue of social desirability bias. Most people want to present a favorable impression of themselves to other people, and this may lead them to distort their answers to personal questions. For example, people are much more willing to admit that they are unhappy when filling in a questionnaire anonymously than when being interviewed (Eysenck, 1990). Second, interviews can only extract information of which the interviewee is consciously aware. This is a significant limitation, because people are often unaware of the reasons why they behave in certain ways (Nisbett & Wilson, 1977). Third, there is the limitation that many interviewers lack some of the skills necessary to conduct interviews successfully. Good interviewers are able to make an interview seem natural, they are sensitive to nonverbal cues, and they have well-developed listening skills (Coolican, 1994).



People are more willing to answer embarrassing or personal questions on an anonymous written questionnaire than in a face-to-face interview.

ETHICAL ISSUES

Interviews (especially clinical interviews) are often concerned with personal issues about which the interviewee is sensitive. This clearly raises the issue of confidentiality. There are various ways in which confidentiality can be broken. For example, Coolican (1994) discussed a study by Vidich and Bensman (1958) in which direct quotations from interviewees in Springdale in the United States were published. Made-up names were used, but the people of Springdale were able to identify the actual individuals on the basis of what they said.

Confidentiality can also be broken if a detailed written account or video recording of an interview falls into the wrong hands. Finally, of course, the interviewer himself or herself may disclose sensitive personal information about the interviewee to other people.

There is another ethical issue that is of particular importance with structured interviews. Interviewees may be aware that several other interviewees are being asked the same questions, and that their answers will be compared. As a result, some interviewees may feel that they must answer embarrassing questions in order not to spoil the experiment.

DISCOURSE ANALYSIS

According to Potter and Wetherell (1987), **discourse analysis** is concerned with “all forms of spoken interaction, formal and informal, and written texts of all kinds.” The basic underlying assumption is that the ways in which we use language are greatly affected by the social context. Thus, for example, when politicians give speeches, it would be naive to assume that what they say simply reflects their genuine beliefs and views. It is generally accepted that what they say is designed to have certain effects on their audience, on other politicians, and on the public.

There is much evidence to indicate that people do adjust what they say or write to fit the circumstances. For example, consider studies using the bogus pipeline. The participants are wired up to an impressive-looking machine (the bogus pipeline), and informed that it can detect any lies they produce. Most white participants express more

Key Term

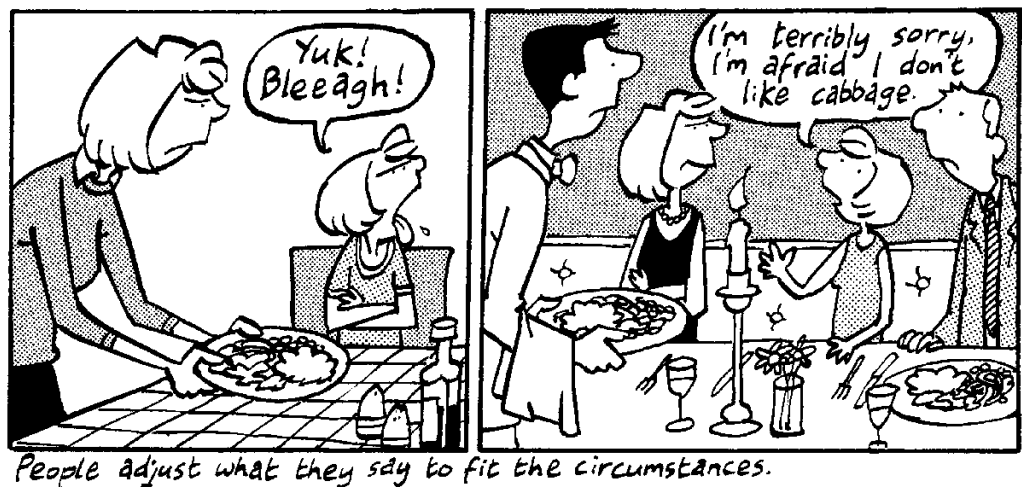
Discourse analysis: a qualitative form of analysis applied to language productions in spoken or written form.

negative attitudes towards black people when wired up to the bogus pipeline than under standard conditions (see Chapter 20). The implication is that the attitudes that people express normally are constructed so as to be socially acceptable to other people.

Gilbert and Mulkay (1984) carried out a discourse analysis based on interviews with 34 scientists. The importance of social factors in discourse was revealed by comparing what these scientists said during the interviews with their academic publications. The general pattern was for scientists to be much more confident about the meaning of their findings when interviewed than they were in their writings.

As Curtis (1997, p. 24) pointed out, “The idea that there is one way to perform discourse analysis is both naive and illusory.” Nevertheless, he identified seven features that are often found in discourse analysis:

1. Select some written or spoken material that is relevant to the issues you want to study.
2. Read or listen to the discourse several times, trying to decide how it has been constructed. Account needs to be taken of the social context in which it was produced.
3. Develop a qualitative coding system focusing on the functions or purposes that seem to be served by the discourse.
4. Produce some tentative hypotheses about the purposes served by the discourse, but be willing to modify these hypotheses if subsequent analysis indicates that they are inadequate.
5. How has the person producing the discourse tried to legitimize or make persuasive his or her version of events?
6. Examine the discourse for evidence of extreme case formulations. People often use extreme terms (e.g., always; never) to make their preferred interpretation seem more persuasive.
7. Examine the discourse carefully to see whether the purposes or functions it serves vary from one part to another.



ADVANTAGES AND LIMITATIONS

One of the advantages of discourse analysis is that it is based on the correct assumption that our use of language is often much influenced by the social context. This is true of how we remember events in our lives as well as our expressed attitudes and beliefs. As Coolican (1994, p.178) pointed out:

When we remember and attribute in real life, as opposed to the psychology experiment, our accounts attend to blame, defense, accountability, explanation and so on. What we often do is to present rememberings as facts when they are really constructions.

For example, the way you describe events in your life is likely to vary depending on whether you are talking to your parents, to your best friend, or to an acquaintance.

Another advantage is that discourse analysis focuses on the ways in which language is used in real-life settings. As such, it avoids much of the artificiality of most experiments. In addition, the claims of those who favor discourse analysis that language is the primary mode of communication among human beings are correct.

There are several limitations of discourse analysis, many of which were discussed by Burman and Parker (1993). A major limitation is that the validity of discourse analysis is open to considerable doubt, and that procedures for assessing validity are lacking. For example, if two researchers interpret a given piece of discourse in very different ways, we cannot be sure which of them has produced the more valid interpretation.

A further limitation is that we often have little information about the reliability or consistency with which the discourse analysis has been carried out. When such information is available, it frequently indicates that reliability is low (Coolican, 1994).

Another limitation is that what emerges from discourse analysis may be unduly influenced by the views and beliefs of the researcher. As Human (1992) expressed it, discourse analysis is sometimes simply “a researcher’s ideas with examples.” A key reason why this can happen is because there are so few constraints on the researcher as he or she tries to make sense of any given written or spoken discourse.

A final limitation is that discourse analysis is based solely on the analysis of language in its various forms. However, language is by no means the only means of communication open to people. Account needs to be taken of nonverbal communication of various kinds (e.g., body language).

ETHICAL ISSUES

It is often important for the researcher to make sure that anyone whose discourse is to be analyzed has given their permission for it to be used for that purpose. However, that ethical issue may not arise if the discourse is in the public domain (e.g., a speech or television interview given by a politician).

There can also be ethical issues if the researcher’s proposed interpretation is likely to offend those who provided the discourse. For example, Wetherell and Potter (1988) carried out discourse analysis on interviews with white New Zealanders. They concluded that those interviewed had racist attitudes towards the Maoris, although they did not directly say so in the interviews. In such circumstances, it is important for the researcher to discuss his or her proposed interpretation with the participants before the results of the study are published or made generally available.

A final ethical issue stems from the fact that discourse analysis often involves detailed analysis of an individual’s discourse. As a result, it is sometimes impossible to adhere to the ethical principles that the information provided by participants should be confidential and that individuals should not be identifiable.

Chapter Summary

Experimental method

- The key principle of the experimental method is that the independent variable is manipulated (with all other variables controlled) in order to observe its effect on some dependent variable. In other words, it is important to avoid confounding variables. The experimental method is used in laboratory and field experiments. Use of the experimental method allows us to infer causality, and it often permits replication. Laboratory experiments have various advantages over field experiments: it is usually easier to eliminate confounding variables, and to obtain detailed behavioral and physiological information. The greatest advantage of field experiments over laboratory experiments is that they are less artificial, and suffer less from factors such as demand characteristics, evaluation apprehension, and the implacable experimenter.

Quasi-experiments

- Quasi-experiments fall short of true experiments either because the experimenter has not manipulated the independent variable or because the participants are not allocated at random to conditions. Natural experiments are quasi-experiments involving some naturally occurring event. Advantages of natural experiments include the possibility that the participants will not be aware they are taking part in an experiment and the opportunity to study the effects of very stressful events. Limitations include problems of interpreting the findings because of a lack of randomization or because of the use of complex independent variables.

Correlational studies

- Correlational designs are inferior to experimental designs because they do not permit inferences about causality. However, many issues can only be studied by assessing correlations or associations between variables. It is often possible to obtain large amounts of data very rapidly in correlational studies. The problems of interpretation are much reduced if there is no correlation or association between two variables.

Naturalistic observation

- Naturalistic observation involves the use of methods designed to assess behavior without the experimenter interfering in any way. Methods of data collection include event sampling, time sampling, and point sampling. We should distinguish between data recording and interpretation or coding. Naturalistic observation can provide rich and full information from people who are unaware that they are being observed. However, the experimenter has essentially no control over the situation, the participants are often aware they are being observed, and there can be problems with the reliability and validity of measurement.

Case studies

- A single individual is investigated thoroughly in a case study. Case studies can be carried out to test a current theory, to refine a theory, to permit the development of new theoretical ideas, and to reveal the exceptional characteristics of certain individuals. Case studies generally have very low reliability, and this makes it hard to generalize from a single-case study. Case studies based on interviews often suffer from the limitation that what the participant says is determined in part by the interviewer or researcher, who may then be too selective in what he or she reports of the interview.

Interviews

- There are several types of interview ranging from the unstructured to the totally structured. Unstructured interviews are responsive to the personality, interests, and motivations of the interviewee, but the data obtained tend to be unreliable. In contrast, structured interviews permit comparisons among interviewees, and they tend to be fairly reliable, but what the interviewee says can be constrained and artificial. All types of interviews can produce problems as a result of social desirability bias, and interviewees can only provide information of which they are consciously aware.

Discourse analysis

- Discourse analysis is based on the assumption that our use of language is much affected by the social context. It involves a careful analysis to identify the underlying purposes of the person who produced the discourse, using a qualitative coding system. Limitations of discourse analysis include low validity and reliability, and the danger that the views of the researcher will influence the findings too much. Ethical issues arise unless the permission of anyone providing discourse for analysis is obtained, and it can be hard to maintain confidentiality of the data.

Further Reading

- Coolican, H. (2004). *Research methods and statistics in psychology* (4th ed.). London: Hodder & Stoughton.
- Dyer, C. (1995). *Beginning research in psychology*. Oxford: Blackwell. This book provides good coverage of several nonexperimental research methods.
- Foster J.J., & Parker, J. (1995). *Carrying out investigations in psychology: Methods and statistics*. Leicester, UK: BPS Books. This is another useful textbook.

chapter 24

Contents

Aims and hypotheses	561
Selecting participants	563
Good practice in experimentation	566
Experimental designs	568
Good practice in nonexperimental designs	570
Problems with experimental research	575
General issues in investigations	578

Design of investigations

24

In order to carry out a study successfully, care and attention must be devoted to each stage in its design and implementation. This chapter is concerned with these issues mostly with respect to experimental designs. However, there is also full consideration of the factors involved in producing good nonexperimental designs.

As we will see, several decisions need to be made when designing an experimental study:

1. The investigator must decide what he or she hopes to achieve by carrying out the study. This involves generating appropriate aims and hypotheses.
2. The investigator has to work out how the variables specified in the hypotheses are to be manipulated and/or measured.
3. Appropriate procedures need to be used when selecting participants for the study.
4. Attention needs to be paid in the experimental design to ensuring that the effects of any situational variables on the participants' behavior are minimized.
5. If the investigator is using an experimental design, then he or she has to select an appropriate one. This includes a decision as to whether each participant will be included in only one condition or in both conditions.
6. Care has to be paid to the relationship between the participants and the investigator in order to prevent systematic biases in the data obtained.

The success or otherwise of the investigator's study can be evaluated in terms of various criteria. If the design and its implementation are appropriate, then the reliability of the findings and their replicability will tend to be high. In addition, use of an appropriate design maximizes the validity of the findings.

AIMS AND HYPOTHESES

The first step that needs to be taken when designing an experimental or nonexperimental study is to decide on the aims and hypotheses of the study. The aims are usually more general than the hypotheses, and they help to explain the reasons for the investigator deciding to test some specific hypothesis or hypotheses. In other words, the aims tell us *why* a given study is being carried out, whereas the hypotheses tell us *what* the study is designed to test.

EXPERIMENTAL STUDIES

The distinction between aims and hypotheses can be seen more clearly if we consider an example. Suppose that we decide to test the levels-of-processing theory put forward by Craik and Lockhart (1972), which states that information that has been processed for meaning will be remembered better than information that has not. In order to do this, we might present all of our participants with the same list of nouns and then ask them to provide free recall 30 minutes later. Half of them might be asked to think of adjectives to go with the nouns (processing of meaning or semantic processing), whereas the other half

What do you hope to achieve with the study?

+

How do you plan to manipulate and measure variables?

+

How will you select participants?

+

Can the effect of situational variables on participants' behavior be minimized?

+

What experimental design will you use?

+

How will bias due to participant/investigator interactions be prevented?

are asked to think of rhyming words (nonsemantic processing). In such a study, the main aim is to investigate levels-of-processing theory. In more general terms, the aim is to see whether long-term memory is influenced by the kind of processing that occurs at the time of learning. The experimental hypothesis is more specific: Free recall from long-term memory is higher when there is semantic processing at the time of learning than when there is non-semantic processing.

NONEXPERIMENTAL STUDIES

The situation with regard to the aims and hypotheses is somewhat different in qualitative research, in which the data are *not* in numerical form. Qualitative research is often based on interviews, observations, or case studies. Qualitative researchers frequently have no specific hypotheses at the outset of the study; rather, the hypotheses to be tested emerge from a detailed consideration of the data. The aims of qualitative research tend to be more general and wide-ranging than those of traditional research.

An example of qualitative research is the work of Marsh, Rosser, and Harré (1978) with soccer fans. Marsh's original aim was to try to understand the aggressive behavior that they often display, but he had few if any preconceptions or hypotheses in mind at the outset of the study. During the course of the study, Marsh et al. (1978) began to realize that there were complex rules or social norms that were shared by football fans, and which played an important role in determining their behavior (see Chapter 19).

HYPOTHESES

Most experimental research starts with someone thinking of an **experimental hypothesis** (also known as the alternative hypothesis). This is simply a prediction or expectation of what will happen in a given situation. For example, you might think of the experimental hypothesis that loud noise will have an effect on people's ability to carry out a task, such as learning the information in a chapter of an introductory psychology textbook.

Variables

As with most experimental hypotheses, the one just mentioned predicts that some aspect of the situation (in this case, the presence of loud noise) will have an effect on the participants' behavior (in this case, their learning of the information in the chapter). In more technical language, the experimental hypothesis refers to an **independent variable**, which is usually some aspect of the experimental situation that is manipulated by the experimenter. In our example, the presence versus absence of loud noise is the independent variable. The hypothesis also refers to a **dependent variable**, which is some aspect of the participants' behavior. In our example, some measure of learning would be used to assess the dependent variable. In a nutshell, most experimental hypotheses predict that a given independent variable will have some specified effect on a given dependent variable.

One-tailed or two-tailed?

It should be noted at this point that there are two types of experimental hypothesis: directional or one-tailed hypotheses, and nondirectional or two-tailed hypotheses. A *directional* or *one-tailed hypothesis* predicts the *nature* of the effect of the independent variable on the dependent variable. In terms of our example, a directional hypothesis might be as follows: Loud noise will reduce people's ability to learn the information contained in the chapter of a textbook. A *nondirectional* or *two-tailed*

Key Terms

Experimental hypothesis: the testable predictions generated by a theory; these usually specify independent and dependent variables.

Independent variable: some aspect of the experimental situation that is manipulated by the experimenter.

Dependent variable: some aspect of the participant's behavior that is measured in a study.

hypothesis predicts that the independent variable will have an effect on the dependent variable, but the *direction* of the effect is not specified. In terms of our example, a nondirectional hypothesis would be as follows: Loud noise will have an effect on people's ability to learn the information contained in the chapter of a textbook. This hypothesis allows for the possibility that loud noise might actually improve learning.

Null hypothesis

The experimental hypothesis consists of the predicted effect of the independent variable on the dependent variable. This can be contrasted with the null hypothesis. The **null hypothesis** simply states that the independent variable will have no effect on the dependent variable. In terms of our example, a suitable null hypothesis would be as follows: Loud noise will have no effect on people's ability to learn the information contained in the chapter of the textbook. In a sense, the purpose of most studies using the experimental method is to decide between the merits of the experimental hypothesis and those of the null hypothesis.

Why do we need a null hypothesis when what we are interested in is the experimental hypothesis? The key reasons are because the null hypothesis is much more precise than the experimental hypothesis, and we need precise hypotheses in order to use statistical tests properly. For example, the null hypothesis that loud noise will have no effect on people's learning ability is precise because it leads to the prediction that the single most likely outcome is that performance will be equal in the loud noise and no noise conditions. Failing that, there will probably only be a small difference between the two conditions, with the difference being equally likely to go in either direction. In contrast, consider the experimental hypothesis that loud noise will reduce people's learning ability. This hypothesis is very imprecise, because it does not indicate how much learning ability will be impaired. This lack of precision makes it impossible to decide the *exact* extent to which the findings support or fail to support the experimental hypothesis.

Manipulating the independent variable

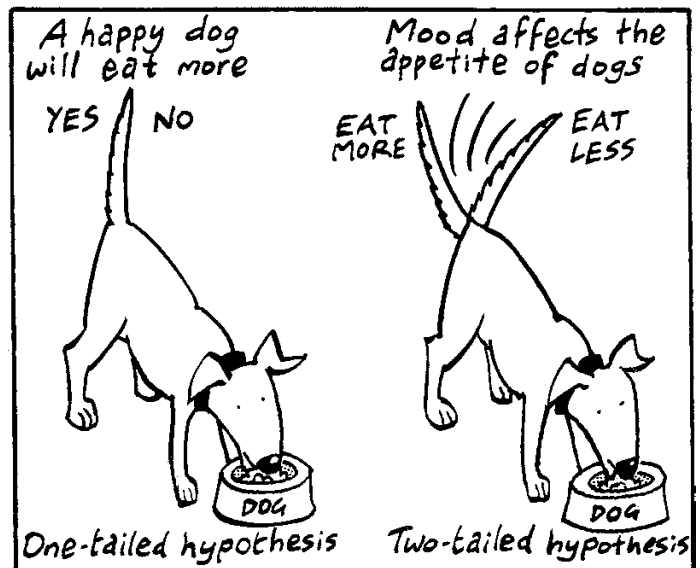
It might seem easy to do a study to test the experimental hypothesis that loud noise disrupts learning. However, there are various pitfalls that need to be avoided. The first issue that needs to be considered is how to manipulate the independent variable. In our example, we want to compare loud noise with no noise, so we have to decide exactly how loud we want the noise to be. If it is very loud, then it might damage the hearing of our participants, and so would be totally unacceptable. If it is fairly soft, then it is unlikely to have any effect on the learning ability of our participants. It is also likely to make a difference whether the noise is meaningful (e.g., music or speech) or meaningless (e.g., the noise of a road drill).

Measuring the dependent variable

The second issue is how to measure the dependent variable or aspect of the participants' behavior. We could ask the participants various questions to measure their understanding of the material in the textbook chapter. However, selecting the questions so that they are not too easy or too hard requires careful thought.

SELECTING PARTICIPANTS

Studies in psychology rarely involve more than about 100 participants. However, researchers generally want their findings to apply to a much larger group of people than those acting as participants. In technical terms, the participants selected for a study form



Key Term

Null hypothesis: prediction that the **independent variable** will have no effect on the **dependent variable**.

a *sample*. This sample is taken from some larger *population*, which consists of all the members of the group from which the sample has been drawn. For example, we might select a sample of 20 children aged 5 for a study, in which case the population might consist of all the 5-year-olds living in one US state.

When we carry out a study, we want the findings obtained from our sample to be true of the population from which they were drawn. In order to achieve this, we must use a *representative sample*, that is, participants who are representative or typical of the population in question. However, numerous studies have been carried out with non-representative samples; the term *sampling bias* is used to refer to this state of affairs. Coolican (1994, p. 36) was pessimistic about the chances of selecting a representative sample:

The simple truth is that a truly representative sample is an abstract ideal unachievable in practice. The practical goal we can set ourselves is to remove as much sampling bias as possible.

RANDOM SAMPLES

To return to our earlier example, we might study the effects of loud noise on learning in students preparing for a psychology exam. The best way of obtaining a representative sample from that population would be to make use of **random sampling**. We could obtain lists of names of all the students due to sit the psychology exam in a given year. After that we could use some random method to select our sample. This could be done by picking names out of a hat, or by sticking a pin repeatedly into the lists.

Another approach is to assign a number to everyone in the population from which the sample is to be selected. After that, a computer can be used to generate a series of random numbers that can be used to select the sample. Alternatively, random number tables can be used in a similar way to produce the sample.

If we wanted to have a representative sample of the entire adult population, then we could apply one of the methods of random selection just described to the electoral roll. However, even that would be an imperfect procedure. Several groups of people, including the homeless, illegal immigrants, and prisoners, are not included in the electoral roll.

As Cardwell, Clark, and Meldrum (1996) pointed out, there is a modified version of random sampling that is easier to use. This is **systematic sampling**. It involves selecting the participants by a quasi-random procedure. For example, if we have a list of all the members of the population, we could select every hundredth name from that list as participants. This procedure is not as effective as random sampling because it cannot be claimed that every member of the population is equally likely to be selected.

Random sampling typically fails to produce a truly representative sample, because it is actually very hard for an experimenter to obtain a random sample. There are various reasons for this. First, it may not be possible to identify all of the members of the larger population from which the sample is to be selected. Second, it may not be possible to contact all those who have been selected randomly to appear in the sample.

Third, some of those who are selected to be in the sample are likely to refuse to take part in the study. This might not matter if those who agreed to take part in research were very similar in every way to those who did not. However, there is considerable evidence that volunteers differ in various ways from nonvolunteers. Manstead and Semin (1996, p. 93) discussed some of the evidence, and concluded, “there *are* systematic personality differences between volunteers and nonvolunteers.” Volunteers tend to be more sensitive to the demand characteristics (cues used by participants to work out what a study is about), and they are also more likely to comply with those demand characteristics.

Systematic sampling is not as effective as random sampling, but it does help to overcome the biases of the researcher. If we select every hundredth name on the list, we avoid missing out names that we cannot pronounce, or do not like the look of.

Key Terms

Random sampling: selecting participants on some random basis (e.g., coin tossing).

Systematic sampling: a modified version of **random sampling** in which the participants are selected in a quasi-random way (e.g., every hundredth name from a population list).

In sum, it is worth bearing in mind what Coolican (1998, p. 720) had to say about random samples: “Many students write that their sample was ‘randomly selected.’ In fact, research samples are very rarely selected at random.”

STRATIFIED AND QUOTA SAMPLES

Another way of obtaining a representative sample is by using what is known as **stratified sampling**. The first step is to decide which characteristics of the population might be relevant for the study we want to carry out. These characteristics might include gender and the part of the country in which they live. This allows us to think in terms of sub-groups. After that, we select participants at random from within each of the sub-groups.

Suppose that we want to carry out a study on A-level psychology students in high schools in England. We know that 75% of A-level psychology students are female, and that 40% of all A-level psychology students live in the north of England. We could then ensure that the participants used in our experiment were selected in a random way so that 75% of them were female, and 40% of them lived in the north of England. If we make use of enough criteria, then stratified sampling can be an effective way of finding a representative sample.

There is a modified version of stratified sampling that is known as **quota sampling**. It resembles stratified sampling in that participants are selected in proportion to their representation in the population. However, it differs in that the researcher decides who to include in each sub-group, rather than the decision being made at random. Quota sampling is often used in market research. It tends to be faster than stratified sampling. However, it has the disadvantage that people who are readily available (e.g., the unemployed) are more likely to be included than those who are not.

The problem with stratified and quota sampling is that it is often hard to know which sub-groups to identify. It is a waste of time and effort if we use characteristics (e.g., gender) that are of no relevance to the study. What is more troublesome is if we fail to identify sub-groups on the basis of some characteristic (e.g., GCSE exam performance in England) that is actually highly relevant.

Stratified sampling is time-consuming and difficult to carry out effectively. Pressures on time and tight budgets may make this sampling technique impossible.

OPPORTUNITY SAMPLING

Random sampling, stratified sampling, and quota sampling are often expensive and time-consuming. As a result, many researchers use **opportunity sampling**. This involves selecting participants on the basis of their availability rather than by any other method. Opportunity sampling is often used by students carrying out experiments, and it is also very common in natural experiments (see Chapter 23). Opportunity sampling is the easiest way to proceed. However, it has the severe disadvantage that the participants may be nothing like a representative sample. For example, students who are friends of the student carrying out a study may be more willing to take part than students who are not.

SAMPLE SIZE

One of the issues that anyone carrying out a piece of research has to consider is the total number of participants to be included. What is the ideal number of participants in each condition? There is no definite answer to that question, but here are some of the relevant factors:

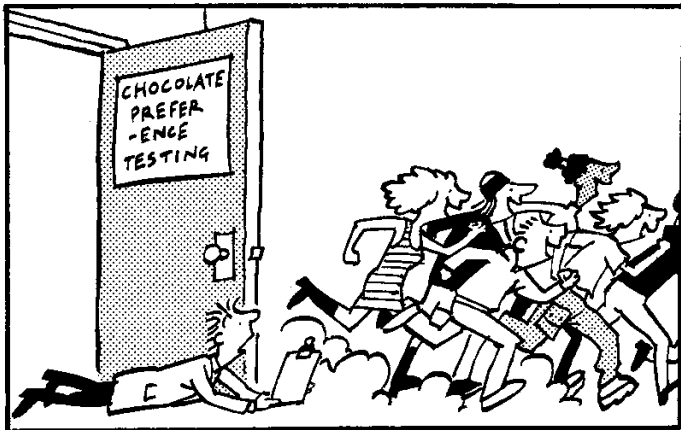
- It is generally expensive and time-consuming to make use of large samples running into hundreds of participants.
- If it requires very large samples to obtain a statistically significant effect of some independent variable on some dependent variable, then this suggests that the effect is small and of little practical importance.
- If we use very small samples (fewer than 10 participants in each condition), then this reduces the chances of obtaining a significant effect.
- In general terms, sampling bias is likely to be greater with small samples than with large ones.

Key Terms

Stratified sampling: a modified version of **quota sampling**, in which the selection of participants according to certain characteristics is decided by the researcher, rather than in a random way.

Quota sampling: selecting participants at random from a population so that they are similar to it in certain respects (e.g., proportion of females; proportion of teenagers).

Opportunity sampling: selecting participants only on the basis of their availability.



Consider the total number of participants to be included...

If there is a golden rule that applies to deciding on sample size, it is the following:

The smaller the likely effect being studied, the larger the sample size needed to demonstrate it.

For most purposes, however, having about 15 participants in each condition is a reasonable number.

GOOD PRACTICE IN EXPERIMENTATION

In order for an experiment to be designed and carried out successfully, there are several considerations that the researcher needs to bear in mind. Some of the main considerations are discussed in detail in this section.

STANDARDIZED PROCEDURES

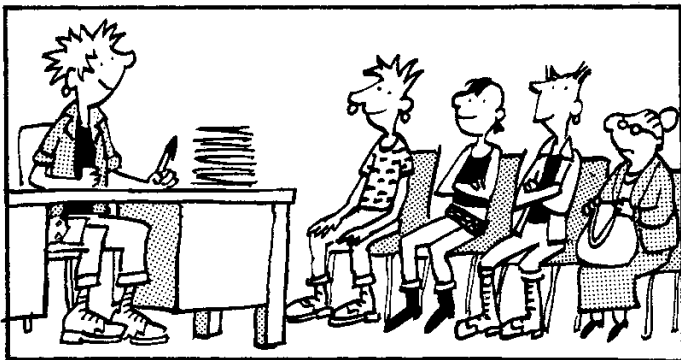
In order to carry out an experiment successfully, it is very important that every participant in a given condition is treated in the same way. In other words, it is necessary to use standardized procedures. For example, consider the instructions that are given to the participants. In order to ensure that all of the participants get precisely the same instructions, the experimenter should write them down. He or she should then either read them to the participants, or ask the participants to read them to themselves.

In similar fashion, standardized procedures should be used for the collection of data. Suppose we want to assess the effects of loud noise on learning from a book chapter. We might ask the participants to write down everything they could remember about the chapter. However, it would be very hard to compare the recalls of different participants with any precision. A standardized procedure would be to ask all of the participants the same set of, say, 20 questions relating to the chapter. Each participant then obtains a score between 0 and 20 as a measure of what he or she has learned.

Is it easy to make sure that standardized procedures are being used? No, it is not. Most experiments can be thought of as social encounters between the experimenter and the participant, and it is customary to behave in different ways towards different people. Robert Rosenthal (1966) studied some of the ways in which experimenters fall short of standardized procedures. He found, for example, that male experimenters were more pleasant, friendly, honest, encouraging, and relaxed when their participants were female than when they were male. This led him to conclude as follows: “Male and female subjects [participants] may, psychologically, simply not be in the same experiment at all.”

Key Term

Confounding variables: variables that are mistakenly manipulated or allowed to vary along with the independent variable.



The type of experimenter could act as a confounding variable. Some participants may feel more comfortable than others in the study situation...

CONFOUNDING VARIABLES

Another issue to consider is whether or not our experiment contains any **confounding variables**. These are variables that are mistakenly manipulated along with the independent variable. Suppose there is a study in which one group of participants receives no noise and reads a chapter at midday, whereas the other group of participants receives loud noise and reads the same chapter at midnight. If we find that the latter group learns less well than the former group, we would not know whether this was because of the loud noise or because they did their learning late at night when they were very tired. In this example, time of day is a confounding variable.

Confounding variables are especially likely to be found in nonexperimental investigations in which the researcher has no control over the independent variable. One of the

classic examples concerns the work on maternal deprivation that was carried out on institutionalized children (see Chapter 16). Bowlby (1951) argued that these children had poorer social and intellectual development than other children because of the absence of the mother. However, these children also had to cope with the unstimulating environment of the institutions of those days, and this was a confounding variable that Bowlby tended to ignore.

Confounding variables are a form of constant error. **Constant error** is present when the effects of any unwanted variable on the dependent variable differ between conditions. There are numerous types of constant error. The participants in one condition may be more tired than those in another condition, or they may be more intelligent, or they may be more motivated.



Any study of the effects of maternal deprivation on these orphans would have to consider the confounding variable of the unstimulating environment of the orphanage itself.

CONTROLLED VARIABLES

How do we avoid having any confounding variables? One useful approach is to turn them into **controlled variables**, which are variables that are held constant or controlled. Suppose that we want to study the effects of noise on learning, and we are concerned that the time of day may have an effect. We could make time of day into a controlled variable by testing all of our participants at a given time of day, such as late morning or early evening. If we did this, we would know that time of day could not distort our findings.

RANDOM ERROR

Random error occurs when variables that are totally irrelevant to the experiment influence the behavior of the participants. The key difference between random error and constant error is that random error generally affects both conditions equally and so has *unsystematic effects*, whereas constant error has *systematic effects* on one condition but not on the other. Constant error is more serious than random error, because it can lead us to misinterpret our findings. However, random error is also of concern, because it introduces unwanted variation in the dependent variable.

There are almost limitless types of random error. For example, suppose we are interested in comparing learning performance under noise and no-noise conditions. Participants in either condition may learn poorly because they have a splitting headache, because they have just argued with a close friend, because a relationship broke up last week, because the weather is bad, or because they are worried about an important examination they have to take next week. The experimenter cannot control most forms of random error, but should try to control those that can be controlled. For example, he or she should ensure that the lighting conditions, the heating conditions, the experimenter's tone of voice, and so on remain constant for all participants.

OPERATIONALIZATION

Psychologists carry out studies to test experimental hypotheses, such as “anxiety impairs performance” or “maternal deprivation leads to maladjustment.” There is an immediate problem with designing a study to test such hypotheses: there is little or no agreement on the best way to measure psychological concepts or variables such as “anxiety,” “performance,” “maternal deprivation,” or “maladjustment.” The most common approach to this problem is to make use of **operationalization**. This involves defining each variable of interest in terms of the operations taken to measure it. Such a definition is referred to as an operational definition. For example, anxiety might be defined as the score on the trait anxiety scale of Spielberger's State-Trait Anxiety Inventory, and performance might be defined as the number of five-letter anagrams that can be solved in 5 minutes.

Key Terms

Constant error:

any unwanted variable that has a systematically different effect on the dependent variable in different conditions.

Controlled variables:

variables, not of interest to the experimenter, that are held constant or controlled.

Random error:

unsystematic and unwanted “nuisance variables” that influence the **dependent variable**.

Operationalization:

a procedure in which variables of interest are defined by the operations taken to measure them.

Operationalization has the great advantage that it generally provides a clear and objective definition of even complex variables. However, there are various limitations associated with the use of operational definitions. First, operational definitions are entirely circular. As Stretch (1994, p. 1076) pointed out:

A psychological construct is defined in terms of the operations necessary to measure it, and the measurements are defined to be measures of the psychological construct.

Second, an operational definition typically only covers part of the meaning of the variable or concept. For example, defining anxiety in terms of the score on a self-report questionnaire largely ignores physiological and behavioral aspects of anxiety, and no one believes that performance can *only* be assessed in terms of rate of anagram solution.

In spite of these important limitations with operational definitions, it is hard to carry out research without using them. Stretch (1994, p. 1076) argued that operational definitions should be used in a careful fashion:

A useful rule of thumb is to consider many different ways of measuring the psychological construct of interest and determine the extent to which each method could yield different experimental results. If you find that the measurement techniques radically affect the results that emerge, this should indicate that more work is needed on developing the underlying psychological and measurement models to explain these effects.

EXPERIMENTAL DESIGNS

If we wish to compare two groups with respect to a given independent variable, it is essential to make sure that the two groups do not differ in any other important way. This general rule is important when it comes to selecting participants to take part in an experiment. Suppose all the least able participants received the loud noise, and all the most able participants received no noise. We would not know whether it was the loud noise or the low ability level of the participants causing poor learning performance.

How should we select our participants so as to avoid this problem? There are three main types of experimental design:

- *Independent design*: Each participant is selected for only one group.
- *Matched participants design*: Each participant is selected for only one group, but the participants in the two groups are matched for some relevant factor or factors (e.g., ability; sex; age).
- *Repeated measures design*: Each participant appears in both groups, so that there are exactly the same participants in each group.

INDEPENDENT DESIGN

With the independent design, the most common way of deciding which participants go into which group is by means of randomization. In our example, this could involve using a random process such as coin tossing to decide whether each participant is exposed to loud noise or to no noise. It is possible with randomization for all the most able participants to be selected for the same group. However, what happens in the great majority of cases is that the participants in the two groups are similar in ability, age, and so on.

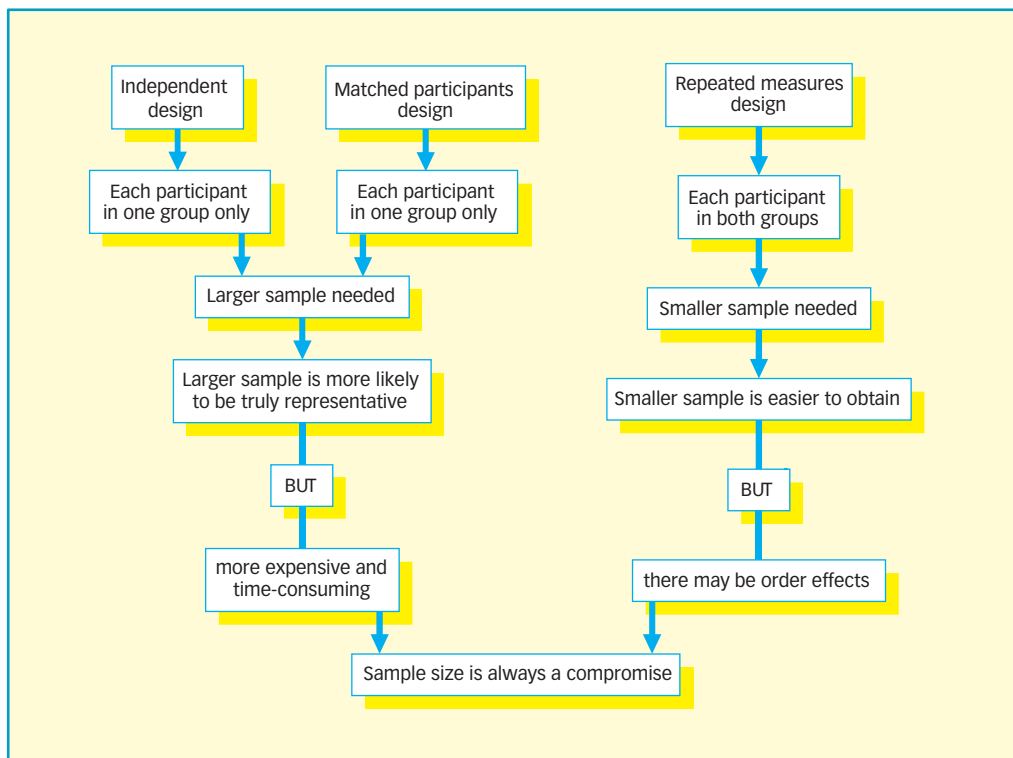
MATCHED PARTICIPANTS DESIGN

With the matched participants design, we make use of information about the participants to decide which group each participant should join. In our example, we might have information about the participants' ability levels. We could then use this information to make sure that the two groups were matched in terms of range of ability.

REPEATED MEASURES DESIGN

With the repeated measures design, every participant is in both groups. In our example, that would mean that each participant learns the chapter in loud noise and that they also learn the chapter in no noise. The great advantage of the repeated measures design is that we do not need to worry about the participants in one group being cleverer than those in the other group: As the same participants appear in both groups, the ability level (and all other individual characteristics) must be identical in the two groups!

The main problem with the repeated measures design is that there may well be order effects. Their experiences during the experiment may change the participants in various ways. They may perform better when they appear in the second group because they have gained useful information about the experiment or about the task. On the other hand, they may perform less well on the second occasion because of tiredness or boredom. It would be hard to use a repeated measures design in our example: Participants are almost certain to show better learning of the chapter the second time they read it, regardless of whether they are exposed to loud noise.

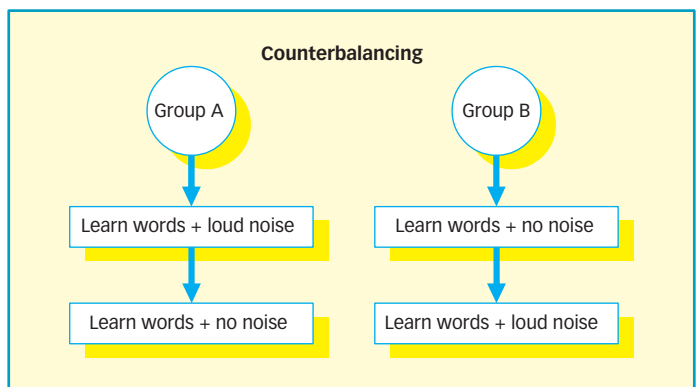


Key Term

Counterbalancing: this is used with the repeated measures design, and involves ensuring that each condition is equally likely to be used first and second by the participants.

Counterbalancing

Suppose we used a repeated measures design in which all of the participants first learned the chapter in loud noise and then learned it in no noise. We would expect the participants to show better learning in no noise simply because of order effects. A better procedure would be to have half the participants learn the chapter first in loud noise and then in no noise, while the other half learn the chapter first in no noise and then in loud noise. In that way, any order effects would be balanced out. This approach is known as **counterbalancing**. It is the best way of preventing order effects from disrupting the findings from an experiment.



GOOD PRACTICE IN NONEXPERIMENTAL DESIGNS

There are several kinds of nonexperimental studies (see Chapter 23). They include naturalistic observation, participant observation, studies based on correlational analysis, interviews and surveys, and case studies. Case studies involve the collection of detailed information from individuals rather than from groups of participants. We begin by considering some general points that need to be taken into account when designing and implementing a nonexperimental study.

GENERAL CONSIDERATIONS

One of the key issues in many nonexperimental studies is to decide whether the participants should be made aware of the fact that they are taking part in research. The main argument for making the participants aware of what is happening is an ethical one. Voluntary informed consent is regarded as of central importance in ensuring that research is ethically acceptable, and it is impossible to obtain that consent from people who do not know they are taking part in a study. However, participants who are made aware may fail to behave in a natural way. Their behavior may be affected by **evaluation apprehension** (desire to impress the experimenter) or by their guesses about the experimental hypothesis being tested (**demand characteristics**).

Some of the issues to be considered are the same as those that apply to experimental studies. For example, if the participants are intended to form a representative sample from a larger population, then it is desirable to use a suitable form of sampling (e.g., random sampling or systematic sampling). The issue of sampling is perhaps especially important with respect to case studies, which involve intensive investigation of individuals. However, there are many nonexperimental studies in which the investigator has little or no control over the selection of participants.

OBSERVATIONAL STUDIES

Observational studies differ from each other in various ways. First, we can distinguish between participant observation research, in which the investigator is involved in the study as an active participant, and nonparticipant observation research, in which the investigator only observes the behavior of the participants.

Second, there is a distinction between unstructured observation and structured observation. According to Dyer (1995, p. 153), unstructured observation is research “where the aim is simply to ensure that everything which appears to be of relevance to the research at a given moment is recorded.” In contrast, an investigator using structured observation makes prior decisions about what to observe, and this “renders the research process relatively inflexible and incapable of responding to unpredictable situations” (1995, p. 154).

Participant observation

The key factor in participant observation is that the researcher has to do his or her best to become accepted by the social group being studied. The goal is to develop a good understanding of what it is like to be a member of that group, and this can only be done when its members accept and trust the researcher. It follows that participant observation research is very time-consuming, because it can take weeks or months for the researcher to gain the confidence of group members.

Dyer (1995) discussed three stages that are involved in carrying out a participant observation study:

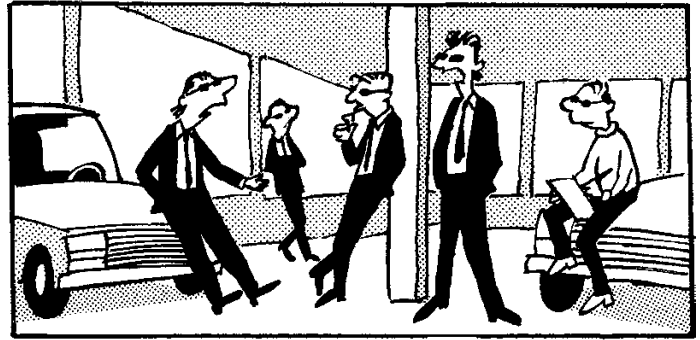
1. *Entering the field*: An important early step is to be accepted by the “gatekeeper” who controls access to the group to be studied; in a school, it is likely to be the Principal. It is usually desirable to let the fact that you are doing research emerge gradually over time. However, there are major ethical issues to be considered, and it is important to have the informed consent of those responsible for the running of the school or other organization.

Key Terms

Evaluation apprehension: anxiety-toned concern felt by participants in a study to perform well and to be evaluated positively by the experimenter.

Demand characteristics: cues used by participants to try to guess the nature of the study or to work out what the experiment is about.

2. *Being in the field:* For the duration of the study, you have the hard task of trying to fit in as a member of the group and of remaining detached as an observer. You should take field notes, which are an extensive record of what members say and do. These field notes should be condensed into a field diary that is written up every day, and should identify key themes. Finally, the field diary is used as the basis for the research report. The initial field notes might include information in the following categories suggested by Lovland (1976): Acts (short actions); activities (actions taking up at least several days); meanings (participants' explanations for their actions); participation (the various roles participants play); relationships among the group members; and settings (the situations in which the group members find themselves).



Try to fit in as a member of the group and remain detached as an observer.

3. *Leaving the field:* There are major ethical issues in participant observation research, because it tends to deal with personal and sensitive issues. It is thus very important to make sure that the group members have the chance to read and comment on your research, and that you take their comments seriously.

Nonparticipant observation

Most nonparticipant observation research starts with the researcher thinking of an experimental hypothesis. If structured observations are to be made, it is then necessary to devise the behavioral categories that are going to be used by the observers. The categories should possess the following features:

1. They should be defined in a reasonably precise and objective way so there is as little ambiguity as possible about which forms of behavior by the participants qualify for each category.
2. The system of categories needs to be comprehensive, in the sense that all aspects of behavior that are relevant to the experimental hypothesis should be included.
3. The categories should be usable in the context of the study. For example, a researcher studying the reactions of drivers stuck in traffic hold-ups might include various categories of facial expression. This is only sensible provided that the observer is going to be able to see drivers' facial expressions clearly from his or her viewing position.

Another key decision concerns the way in which the participants' behavior is to be sampled. Dyer (1995) identified four possible sampling procedures:

1. *Continuous observation:* The observer records behavior falling into the various categories nonstop over a fairly lengthy period of time (e.g., 60 minutes).
2. *Time-interval sampling:* The sampling period is divided into a series of short time intervals (e.g., 60 seconds), and the observer decides whether any given participant produces each category of behavior during each period. Any behavior is simply recorded as absent or present, so that no distinction is drawn between a given behavior exhibited once versus more than once during a single time interval.
3. *Time-point sampling:* The sampling period is divided into a series of short time intervals, and the observer decides whether the various categories of behavior are present at the end of each sampling period.
4. *Random sampling:* This is like time-point sampling, except that the points in time at which behavior is sampled are selected at random.

SURVEY STUDIES

The survey method involves collecting information from a large group of individuals. This information is often gathered using questionnaires, but can include interviews or phone contacts. It is important in any survey study to ensure that the sample selected is as representative as possible (see earlier). A problem that applies to nearly all sampling

methods is that of nonresponding. Some individuals who are selected to form part of the sample are likely to refuse to participate. Others agree to participate, but then fail to provide all of the information requested. Persuasion and persistence should be used to minimize the problem of nonresponding, but it is very rare for the response rate to be 100% in a survey study.

Survey designs

According to Dyer (1995), there are four main types of survey: one-shot survey; before–after design; two-groups controlled comparison design; and two-groups before–after design.

One-shot surveys. The one-shot survey is the simplest, but also generally the least informative type of survey. Information is obtained from a single sample at a given point in time. The reason why it is fairly uninformative is that we cannot compare the findings from our sample against those of other groups. As a result, we can only describe what we have found to be the case in the sample we tested.

Before–after surveys. The before–after design is an advance on the one-shot survey, in that data are collected from a single sample on two occasions. The design is most likely to produce interesting findings if some major event or experience intervenes between the first and second data collections. For example, attitudes in the UK towards the Labour party could have been obtained shortly before and after the general election of 1997. Suppose (as seems to have been the case) that attitudes were more positive after the election than before. This might have been due to the election victory, but there are other possibilities. Attitudes to the Labour party might have become more positive even if there had not been an election, or it may be that people tend to respond differently on the second occasion that an attitude questionnaire is completed than on the first. In general, it is hard to interpret the findings based on the before–after design.

Two-groups controlled comparison surveys. The two-groups controlled comparison design is potentially more informative than the designs discussed so far. In essence, there are two similar groups of participants, one of which is exposed to some treatment before data collection, whereas the other is not. For example, attitudes towards the opposite sex could be assessed in those who have (or have not) recently experienced the breakdown of a heterosexual relationship. If the former group was more negative in their attitudes, it could be argued that this was due to the breakdown of the relationship. However, this requires the assumption that the two groups had the same attitudes before the breakdown occurred, and we cannot be sure that that assumption is justified.

Two-groups before–after surveys. The two-groups before–after design is an advance on the two-groups controlled comparison design. Two samples or groups are tested for the first time, then one group is exposed to some treatment, and finally both groups are tested for a second time. Dyer (1995) gave as an imaginary example a study in which the participants are allocated at random to two groups. The attitudes of all of them towards Third World issues are assessed. One group is then exposed to several presentations of a television commercial focusing on the need to provide economic aid to Third World countries. Finally, the attitudes of both groups towards Third World countries are assessed. This survey method is the most complicated one to use, but the findings are easier to interpret than those from other survey methods.



Do events such as a general election victory affect attitudes towards a political standpoint, or vice versa?

Questionnaire construction

In order to address the specific issues that interest them, researchers using the survey method often construct their own questionnaire. The first step is to generate as many ideas as possible that might be relevant to the questionnaire. Then those ideas that seem of little relevance are discarded, working on the basis (Dyer, 1995, p. 114) that:

It is better to ask carefully designed and quite detailed questions about a few precisely defined issues than the same number on a very wide range of topics.

Closed and open questions. There is an important distinction between closed and open questions. Closed questions invite the respondent to select from various possible answers (e.g., yes or no; yes, unsure, or no; putting different answers in rank order), whereas open questions allow respondents to answer in whatever way they prefer. Most questionnaires use closed questions, because the answers are easy to score and to analyze. Open questions have the disadvantage of being much harder to analyze, but they can be more informative than closed questions.

Question styles: A survey on chocolate

Closed question: Do you like chocolate? (tick one)

☐ YES ☐ NO ☐ NOT SURE

Open question: Why do you like or dislike chocolate?

Ambiguous question: Is chocolate likely to do you more harm than a diet that consists mainly of junk food?

Biased question: Plain chocolate is a more sophisticated taste than milk chocolate. Which type do you prefer?

Ambiguity and bias. Questions that are ambiguous or are likely to be interpreted in various ways should be avoided. Questions that are very long or complicated should also be avoided, because they are likely to be misunderstood. Finally, questions that are biased should be avoided. Here is an example of a biased question: “In view of the superiority of Britain, why should we consider further political integration with the rest of Europe?”

Reliability and validity. Good questionnaires need to have high reliability or consistency of measurement. They also need reasonable validity, meaning that they should measure what they claim to measure. Reliability can be assessed by means of the test–retest method, in which the questionnaire is given to the same individuals on two different occasions. The scores can then be correlated by means of a test such as Spearman’s rho (see Chapter 25). If the correlation is fairly high (about +0.7 or +0.8), then the questionnaire can be regarded as reliable.

There are several ways of assessing the validity of a test. For example, there is empirical validity, in which the scores on a questionnaire are compared against some external criterion. For example, suppose that someone devised a questionnaire to measure conscientiousness. It seems reasonable to assume that conscientious people will perform better on examinations, and so we could use examination performance as the external criterion. Conscientiousness scores on the questionnaire could be correlated with examination performance using Spearman’s rho, with the assumption being that there would be a significant positive correlation.

Attitude scale construction

Many of the points made about questionnaire construction also apply to the construction of attitude scales. However, there are some differences.

Likert scales. One of the most common ways to construct an attitude scale is to use the Likert procedure. Initially various statements are collected together, and the participants’ task is to indicate their level of agreement on a five-point scale running from “strongly disagree” at one end to “strongly agree” at the other end. For positive statements (e.g., “Most Hollywood stars are outstanding actors”), strongly disagree is scored as 1 and strongly agree as 5, with intermediate points being scored 2, 3, or 4. For negative statements (e.g., “Most Hollywood stars are not outstanding actors”), the scoring is reversed so that strongly disagree is scored as 5 and strongly agree as 1.

Most attitude scales based on the Likert method contain some unsatisfactory items. One way of finding out which items are unsuitable is by correlating each item separately with the total score on the scale. Only items that correlate positively at a moderate level with the total score (+0.3 and above) are retained for the scale.

Reliability and validity. The reliability of an attitude scale can be assessed by the test–retest method. Its validity can generally be assessed by some measure of empirical validity. For example, we could obtain evidence about the validity of a scale concerned with attitudes towards religion by correlating the scores with a measure such as regularity of attendance at church, by using Spearman’s rho. However, it is important

to note that the correlation may be low either because an attitude scale lacks validity or because there is often a large difference between people's attitudes and their behavior (see Chapter 19).

CORRELATIONAL STUDIES

Correlational studies typically involve obtaining two different measures from a group of participants, and then assessing the degree of association between the measures by using a test of correlation such as Spearman's rho. For example, participants' level of extraversion could be correlated with their number of friends, based on the prediction that extraverts are likely to have more friends than introverts.

Correlational studies are easy to carry out. For example, there are thousands of questionnaires for measuring personality or attitudes, and it is possible to take any two at random and administer them to a large group of people. After that, the scores on the two questionnaires can be correlated. However, the fact that correlational studies are easy to perform does *not* mean that good correlational studies are easily carried out. What features characterize good correlational studies?



Correlation or causation?

An underlying theory

First, the study should be based on some underlying theory. The two variables that are measured in the study should both be of clear relevance to the theory. In addition, the predicted direction of the correlation (positive or negative) should follow from the theory. For example, there is the matching hypothesis, according to which we are attracted to those who are about as physically attractive as we are. This was tested in a correlational study by Murstein (1972). The physical attractiveness of couples was judged from photographs. There was a strong positive correlation, with the most physically attractive individuals tending to have someone very attractive as their partner, whereas those who were physically unattractive had unattractive partners.

In many correlational studies, one of the variables can be regarded as the predictor variable with the other one as the outcome variable. The predictor variable can be seen as

occurring before the outcome variable in some sense. It is called the predictor variable because it forms the basis for predicting the value of the outcome variable. For example, there is a positive correlation between the Type A behavior pattern (hostility, impatience, tension) and coronary heart disease (Miller et al., 1991). Here the Type A behavior pattern is the predictor variable and coronary heart disease is the outcome variable. This approach may suggest the existence of a causal relationship. However, it is very important to remember that “correlations cannot prove causes.” There could be a third factor (e.g., genetic vulnerability) that leads to the Type A behavior pattern and to susceptibility to heart disease.

Careful measurement

Another feature of good correlational studies is that the variables are carefully measured. Let us consider an example. Martin, Kulper, and Westra (1989) argued that Type A individuals are much more highly motivated than Type B individuals, who are relaxed, patient, and calm. This suggests that Type A individuals might have better job performance than Type Bs; thus, there should be a positive correlation between Type A and job performance. How can job performance be measured? In the case of managers whose jobs involve forward planning, motivating their staff, monitoring the performance of their staff, and so on, it may be very hard to assess their work performance in a single measure. It would be preferable to study a group such as

insurance salespeople. Their main work goal is to sell as much insurance as possible, and so the amount of insurance sold over a given period of time (e.g., 3 months) would provide a reasonable measure of job performance.

Wide range

A final feature of good correlational studies is that the scores on both variables vary considerably from individual to individual. For example, IQ is supposed to reflect general intellectual ability, and so one might predict that there would be a positive correlation between IQ and job performance. This has been found numerous times (Eysenck, 1994). Suppose, however, that we correlated IQ and job performance among chartered accountants. The great majority of chartered accountants have high IQs, and so we have what is known as restriction of range. This restriction of range would reduce the strength of the association between IQ and job performance, and so it should be avoided.

PROBLEMS WITH EXPERIMENTAL RESEARCH

In most experimental research (and some nonexperimental research), the experimenter and the participants interact with each other. This can produce various kinds of problems. The ways in which experimenters behave and talk may influence the behavior of the participants in ways that have nothing to do with the independent variable or variables being manipulated. In addition, the participants may form mistaken ideas of what the experiment is about, and these mistaken ideas may affect their behavior. Some of the main problems stemming from the relationship between the researcher and the participants are discussed in this section.

EXPERIMENTER EFFECTS

The ideal experimenter is someone who behaves in exactly the same mildly positive way with every participant, and who does not allow his or her expectations and experimental hypotheses to influence the conduct of a study. In reality, the experimenter's expectations, personal characteristics, and so on often have an effect on the participants' behavior; these are known as **experimenter effects**.

Experimenter expectancy

One of the most important experimenter effects is experimenter expectancy, in which the experimenter's expectations have a systematic effect on the performance of the participants. Perhaps the first systematic demonstration of experimenter expectancy involved a horse known as Clever Hans. The horse was apparently able to count, tapping its hoof the right number of times when asked a simple mathematical question (e.g., $8 + 6$). Pfungst (1911) studied Clever Hans. He found that Clever Hans could not produce the correct answer when the horse was blindfolded. What happened normally was that the experimenter made slight movements when the horse had tapped out the correct number, and Clever Hans was simply using these movements as the cue to stop tapping.



The way in which experimenters behave and talk may influence the behavior of the participant.

Key Term

Experimenter effects:

the various ways in which the experimenters' expectancies, personal characteristics, misrecordings of data, and so on can influence the findings of a study.



Clever Hans, the "counting" horse.

Key Study

Rosenthal (1966): Experimenter effects

One of the best-known studies on experimenter effects was reported by Rosenthal (1966). He asked student experimenters to count the number of head turns and body contractions made by flatworms. Before the experiment started, the students were told that they should expect a lot of activity from half of the worms, but very little activity from the others. In fact, worms were assigned at random to the two groups, so there was no reason for assuming that they would actually differ in activity level.

What do you think Rosenthal found? Somewhat surprisingly, the experimenters reported *twice* as many head turns and *three* times as many body contractions in the worms that were allegedly “highly active” as in the “inactive” ones! Rosenthal argued that this was an experimenter expectancy effect, but it is more likely that it was a result of the experimenters failing to follow the proper procedures and/or misrecording of the data. As Coolican (1994) pointed out, there was no evidence of expectancy effect in at least 40 experiments specifically designed to find it. There is evidence that the behavior of human participants, especially those high in need for approval, can be influenced by the experimenter’s behavior. However, it seems less likely that flatworms would respond to a smile or a frown from the experimenter!

Discussion points

1. Are you surprised that it has proved hard to replicate Rosenthal’s findings on flatworms?
2. In what circumstances would you expect to find experimenter effects (see later)?

KEY STUDY EVALUATION

Psychological experiments like Rosenthal’s are carried out by humans on humans. As such they are unique social situations in which social interactions play an important part. Inevitably, problems can arise in the form of experimenter effects. According to Rosenthal, the participants could be influenced to expect certain results to occur within the experiment. However, it is possible that the experimenter could have given clues as to how the participants were expected to behave, either verbal or nonverbal in nature. Rosenthal suggested that the perceived competence and authority of the research could also produce experimenter effects. This would be influenced by the participant’s own personal characteristics such as a need for approval.

Other effects

Barber (1976) argued that there are numerous ways in which the experimenter can influence the findings obtained. In addition to experimenter expectancy, he identified several other kinds of experimenter effects (summarized in Coolican, 1994). These effects are listed here. In this list, a distinction is drawn between the investigator (the person *directing* the research) and the experimenter (the person actually *carrying out* the experiment). For example, an academic psychologist will often be the investigator, whereas an undergraduate or postgraduate student is the experimenter. So far as the studies carried out by undergraduate studies are concerned, the investigator and the experimenter will typically be the same person.

1. *Investigator paradigm effect*: The entire approach adopted by the investigator can make it harder or easier to obtain certain findings.
2. *Investigator experimental design effect*: For example, if an investigator wanted to show that a learning program for disadvantaged children was not effective, he or she

could arrange for the program to last for a very short period of time to reduce the chances that there would be any noticeable effects.

3. *Investigator loose procedure effect*: If the instructions and other aspects of the procedure are not clearly specified, there is more scope for the results to be influenced by the investigator.
4. *Investigator data analysis effect*: The investigator can decide to carry out several unplanned analyses of the data *after* seeing what patterns seem to be present.
5. *Investigator fudging effect*: There is evidence that Burt, who believed strongly that intelligence depends on heredity, fudged some of his twin data.
6. *Experimenter personal attributes effect*: For example, an experimenter who liked women but not men might treat male and female participants differently, and so produce a spurious gender effect in the data.
7. *Experimenter failure to follow the procedure effect*: If this happens, then the independent variable may not be manipulated as it should be, which makes it hard to interpret the findings.
8. *Experimenter misrecording effect*: Experimenters are most likely to misrecord data if the information provided by participants is ambiguous (e.g., did the participant give a little smile?).
9. *Experimenter fudging effect*: The experimenter may fudge the data to please the investigator or to obtain good marks for his or her study.

Reducing experimenter effects

What steps can be taken to minimize experimenter effects? One approach is to use a **double blind** procedure, in which neither the experimenter working with the participants nor the participants knows the experimental hypothesis (or hypotheses) being tested. The double blind procedure reduces the possibility of experimenter bias, but it is often too expensive and impractical to use. However, the incidence of experimenter effects is probably less than it used to be, for the simple reason that more and more experiments involve participants interacting with computers rather than with human experimenters. In addition, data are increasingly stored directly in computers, making it harder to misrecord the information obtained from participants.

DEMAND CHARACTERISTICS

A common criticism of laboratory research is that the situation is so artificial that participants behave very differently from the way they do normally. Guy Claxton (1980) discussed an amusing example of this. He considered a laboratory task, in which participants have to decide as rapidly as possible whether sentences such as “Can canaries fly?” are true or false. Under laboratory conditions, people perform this task uncomplainingly. However, as Claxton pointed out, “If someone asks me ‘Can canaries fly?’ in the pub, I will suspect either that he is an idiot or that he is about to tell me a joke.”

Why do people behave in unusual ways under laboratory conditions? The American psychologist Orne (1962) emphasized the importance of what he termed *demand characteristics*, which are “the totality of cues which convey an experimental hypothesis to the subjects [participants].” Demand characteristics include “the rumors or campus scuttlebutt about the research, the information conveyed during the original situation, the person of the experimenter, and the setting of the laboratory, as well as all explicit and implicit communications during the experiment proper.” (In case you are wondering, the word “scuttlebutt” means gossip.) Orne’s basic idea is that most participants do their best to comply with what they perceive to be the demands of the experimental situation, but their perception will often be inaccurate.

As Orne showed, the demand characteristics in an experiment are so powerful that the participants can often be persuaded to do some very strange things. He discussed one study in which the participants spent several hours adding numbers on random number sheets, then tearing up each completed sheet into at least 32 pieces. Many of the participants treated the situation as a test of endurance, and this motivated them to keep going.

Key Term

Double blind:

a procedure in which neither the experimenter nor the participants knows the precise aims of the study; where possible, they do not know the condition in which each participant has been placed.



There is another problem with demand characteristics, which applies to participants who have previously taken part in an experiment in which they were deceived about the experimental purpose. As a result of being deceived, some participants tend thereafter to respond in the opposite direction to the one suggested by an experiment's demand characteristics. Why should this be so? Silverman, Shulman, and Wiesensthal (1970) explained this effect in the following way:

Deceived subjects [participants] may have become so alerted to possible further deceptions that they tend to respond counter to any cues regarding the experimenter's hypothesis. An element of gamesmanship may enter the experimental situation in that subjects [participants] become wary of "tricks" underlying the obvious, and do not want to be caught in them.

Reducing demand characteristics

Information about the demand characteristics in any given experimental setting can be obtained by asking the participants afterwards to describe in detail what they felt the experiment was about. Armed with this information, the experimenter can take steps to make sure that the results of future experiments are not adversely affected by demand characteristics.

Some (but not all) of the problems of demand characteristics can be reduced by the double blind procedure described earlier. Another possibility in some studies is the **single blind** procedure, in which the participants are not informed of the condition in which they have been placed. However, this raises ethical issues, because full informed consent cannot be obtained in such circumstances.

EVALUATION APPREHENSION

Demand characteristics are not the only reason why we might expect participants in experiments to behave in ways that differ from their typical everyday behavior. As we saw in Chapter 23, Rosenberg (1965) argued that it is important to take account of what he called *evaluation apprehension*. What this means is that most participants in experiments are concerned to be evaluated positively by the experimenter or, failing that, to avoid being evaluated negatively. It could be argued that the main reason why participants comply with the demand characteristics of experimental situations is because of their evaluation apprehension. However, that would be overstating matters. As discussed in Chapter 23, Sigall, Aronson, and Van Hoose (1970) set up a situation in which evaluation apprehension would lead participants to perform a task rapidly but demand characteristics would lead them to perform it slowly. What they found was that the participants' performance was determined more by evaluation apprehension than by demand characteristics.

GENERAL ISSUES IN INVESTIGATIONS

So far in this chapter, we have considered several specific issues that are important to ensure that the design of a study is appropriate. In this section, we address some important general criteria that can (and should) be used to evaluate how successfully a study has been designed and carried out. The criteria to be discussed are participant reactivity; validity; generalizability; and reliability.

PARTICIPANT REACTIVITY

A weakness that is found in many studies is what is known as **participant reactivity**. This refers to a situation in which an independent variable has an effect on behavior simply because the participants know that they are being observed or studied. Any measure of the participants' behavior that could suffer from this effect is called a reactive measure, and reactivity is the term used to refer to the changes in behavior produced in this way.

Key Terms

Single blind:

a procedure in which the participants are not informed of the condition in which they have been placed.

Participant reactivity:

the situation in which an independent variable has an effect on behavior merely because the participants are aware they are being observed.

Key Study

Roethlisberger and Dickson (1939): The Hawthorne effect

In order to clarify the meaning of participant reactivity, we will consider a series of studies carried out at the Hawthorne Western Electric plant in Chicago (Roethlisberger & Dickson, 1939). They found that the workers became more productive when the amount of lighting was increased, suggesting that work rate increases when the working conditions become easier. However, they also found that *decreasing* the amount of lighting led to increased productivity! In general, it was found that productivity increased when *any* changes were made to the working conditions, whether these changes were to wages, length of the working day, or to rest. Productivity even improved when there was a return to the original working conditions. Presumably what was happening was that the workers responded to the interest being shown in them, rather than to the specific changes in their working environment.

The term “Hawthorne effect” came to be used to refer to changes produced because people know they are being studied, although the same phenomenon is now generally referred to as participant reactivity. There are several published findings that may have been influenced by participant reactivity. For example, Klaus and Kennell (1976) reported that mothers who were allowed to interact for several hours a day with their newborn babies developed a closer relationship with them than did mothers who spent less time with their babies (see Chapter 16). The extra-contact mothers were mostly unmarried teenagers, and it has been argued that this effect was a result of the interest shown in them by the hospital workers rather than to the extra contact itself. This interpretation is supported by the fact that this finding has generally not been replicated in studies of mothers who are less likely to be flattered at being the center of attention (Durkin, 1995).

Participant reactivity or the Hawthorne effect is a serious problem, because it can lead us to misinterpret our findings. How can we decide whether some effect is a result of participant reactivity? In essence, we need to make sure that participant reactivity is the same in both conditions, by making it equally clear to both groups that they are being studied and that their behavior is of interest. If the effect is still found, then it cannot have been caused by participant reactivity. For example, if extra contact of mothers and babies was still associated with a closer relationship even when equal interest was shown in the mothers given only routine contact, then it would be reasonable to conclude that it was the contact itself that produced the effect.

Discussion points

1. How much of a problem is participant reactivity in research?
2. When would you *not* expect to find evidence of participant reactivity?

KEY STUDY EVALUATION

Examining human relations in the workplace grew out of Roethlisberger and Dickson’s (1939) study, which aimed to consider the relationship between working conditions and productivity. The initial emphasis was on the extrinsic rewards the worker received, and it was found that there was no relationship between extrinsic rewards and productivity. What became apparent was that intrinsic rewards had a greater effect. These intrinsic rewards derived from the workers’ own attitudes towards their work both individually and as part of an informal group. The human need to be part of a social group and to be accepted within it determines attitudes to work and the motivation needed to perform successfully far more than financial rewards. From the Hawthorne study new research was stimulated, which examined the range of needs experienced by the workforce. To increase productivity it was found that social needs had to be met, such as friendship, group support, acceptance, approval, recognition, status, and the need for “self-actualization,” which involves the development of an individual’s talents, creativity, and personality to the full.

VALIDITY AND GENERALIZABILITY

One of the key requirements of a study or experiment is that any findings obtained are valid, in the sense that they are genuine and provide us with useful information about the phenomenon being studied. Campbell and Stanley (1966) drew a distinction between internal validity and external validity, which is of most relevance to experiments and quasi-experiments. **Internal validity** refers to the issue of whether the effects observed are genuine and are caused by the independent variable. In contrast, **external validity** refers to the extent to which the findings of a study can be generalized to situations and samples other than those used in the study. This distinction between two kinds of validity is an important one: many experiments possess internal validity while lacking external validity (see Chapter 23).

Internal validity

We will shortly consider some of the reasons why an experiment may lack external validity, but what are some of the main threats to the internal validity of an experiment? Coolican (1994) pointed out that there are many such threats, most of which were discussed earlier in the chapter. For example, the existence of any confounding factors threatens internal validity, as does the use of unreliable or inconsistent measures. Problems with internal validity can also arise if an experiment is designed without careful attention being paid to issues such as standardization, counterbalancing, and randomization. Other threats to internal validity include experimenter effects, demand characteristics, participant reactivity, and the use of inappropriate statistical tests. In a nutshell, virtually all of the principles of experimental design are intended to enhance internal validity, and failures to apply these principles threaten internal validity. If internal validity is high, then there are good prospects for being able to replicate the findings. If it is low, then replication is likely to be difficult or impossible.

External validity and generalizability

What about external validity? There are close links between external validity and **generalizability**, because both are concerned with the issue of whether the findings of an experiment or study are applicable to other situations. More specifically, Coolican (1994) argued that there are four main aspects to external validity or generalizability, which we consider in turn:

- **Populations:** Do the findings obtained from a given sample of individuals generalize to a larger population from which the sample was selected?
- **Locations:** Do the findings of the study generalize to other settings or situations? If the findings generalize to various real-life settings, then the study is said to possess ecological validity. Silverman (1977, p. 108) was skeptical about the ecological validity of laboratory experiments: “the conclusions we draw from our laboratory studies pertain to the behavior of organisms in conditions of their own confinement and control and are probably generalizable only to similar situations (institutions, perhaps, such as schools or prisons or hospitals).”
- **Measures or constructs:** Do the findings of the experiment or study generalize to other measures of the variables used? For example, suppose we find that people who are high on the personality dimension of trait anxiety as assessed by Spielberger’s State–Trait Anxiety Inventory have worse long-term memory measured by recall than those low in trait anxiety. Would we obtain the same findings if trait anxiety were assessed by a different questionnaire or if we used a recognition test of long-term memory?
- **Times:** Do the findings generalize to the past and to the future? For example, it could be argued that the sweeping changes in many cultures in recent decades have affected conformity behavior as studied by Asch, and obedience to authority as studied by Milgram (see Chapter 19).

What can we do to maximize the external validity of an experiment? Unfortunately, there is no easy answer to that question. What usually happens is that the external validity of an experiment only becomes clear when other researchers try to generalize the findings to other samples or populations, locations, measures, and times. It might be thought that the findings of field experiments are more likely than those of laboratory experiments to generalize to other real-life locations or settings, but that is not necessarily so.

Key Terms

Internal validity:

the validity of an experiment in terms of the context in which it is carried out, including the extent to which its findings can be replicated; also the extent to which research findings are genuine and can be regarded as being caused by the **independent variable**; see **external validity**.

External validity:

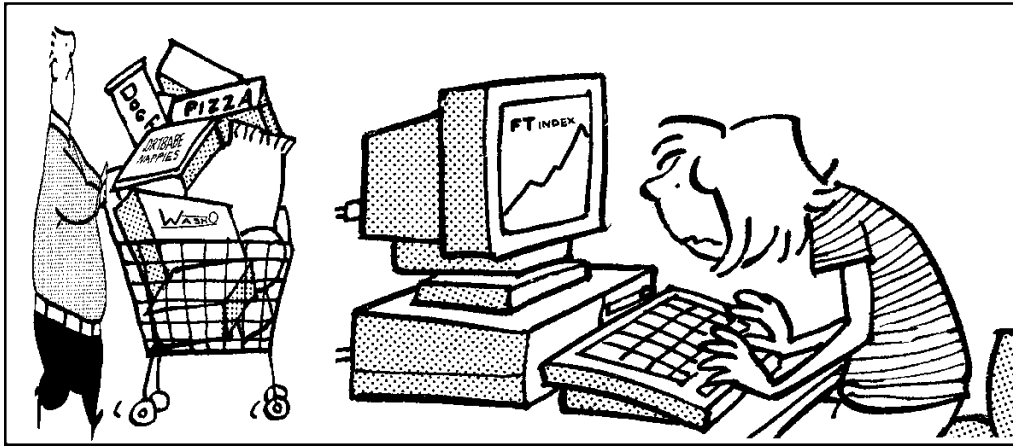
the validity of an experiment outside the research situation itself; the extent to which its findings are applicable to everyday life and generalize across populations, locations, measures, and times; see **internal validity**.

Generalizability:

the extent to which the findings of a study apply to other settings, populations, times, and measures.



Surveys of women’s daily activities and attitudes towards domestic work carried out even a few decades ago bear little relevance to conditions and attitudes current today.



Meta-analyses. One way of trying to determine whether certain findings generalize is to carry out what is known as a **meta-analysis**. What is done in a meta-analysis is to combine all of the findings from many studies designed to test a given hypothesis into a single analysis. If the meta-analysis indicates that some finding has been obtained consistently, this suggests that it generalizes across populations, locations, measures, and times. For example, Smith, Glass, and Miller (1980) discussed a meta-analysis on over 400 studies concerned with the effectiveness of psychotherapy. They concluded that psychotherapy was reasonably effective, because patients receiving psychotherapy improved more than did 75% of the patients not receiving any therapy.

The greatest limitation of meta-analyses is that differences in the quality of individual studies are often ignored. This can lead to the situation in which a finding is accepted as genuine when it has been obtained in several poorly designed studies but not in a smaller number of well-designed studies. Another problem is that it is often hard to know which studies to include and which to exclude. For example, Smith et al. considered all forms of nonbehavioral therapy together. However, some forms of nonbehavioral therapy were more effective than others (Barlow & Durand, 1995), so it was perhaps undesirable to put them together into a single meta-analysis.

Key Term

meta-analysis:

an analysis in which all of the findings from many studies relating to a given hypothesis are combined for statistical testing to obtain an overall picture.

RELIABILITY

One of the main goals of experimental research is to design and carry out studies in such a way that *replication* or repetition of its findings is possible. In order to achieve that goal, it is important that the measures we use should possess good *reliability* or consistency. As Coolican (1994, p. 50) pointed out:

Any measure we use in life should be reliable, otherwise it's useless. You wouldn't want your car speedometer or a thermometer to give you different readings for the same values on different occasions. This applies to psychological measures as much as any other.

Problems relating to reliability are likely to arise when the experimenter is trying to code the complex behavior of participants using a manageable number of categories. For example, it is common in studies of naturalistic observation to record certain events (e.g., performing an aggressive act). However, it may be hard to define those events with enough precision to produce reliable results. One way of assessing this is by asking two (or more) judges to provide ratings in the observational situation. The ratings can then be compared to provide a measure of interjudge reliability.

Reliability



Internal reliability = consistency within the method of measurement. For instance, a ruler should be measuring the same distance between 0 and 5 centimeters as between 5 and 10 centimeters.

External reliability = consistency between uses of the method of measurement. For instance, the ruler should measure the same on a Monday as it does on a Friday.

- Reliability = consistent and stable
- Validity = measuring what is intended
- Standardization = comparisons can be made between studies and samples

Chapter Summary

Aims and hypotheses

- The first stage in designing a study is to decide on its aims and hypotheses. There will generally be an experimental or alternative hypothesis and a null hypothesis. The experimental hypothesis may be directional or one-tailed, or it may be nondirectional and two-tailed.

Selecting participants

- The participants selected for a study represent a sample from some population. They should form a representative sample; in other words, sampling bias should be avoided. The best approach is random sampling, but other reasonable methods are systematic sampling, stratified sampling, and quota sampling. Opportunity sampling is the easiest but least satisfactory method. The sample size depends on the likely size of the effect being studied.

Good practice in experimentation

- It is important to use standardized procedures. It is also important to avoid confounding variables and other forms of constant error, and to keep random error to a minimum. Operationalization is useful, but operational definitions typically cover only part of the meaning of the independent or dependent variable in question.

Experimental designs

- There are three main types of experimental design: independent design; matched participants design; and repeated measures design. With an independent design, randomization is generally used to allocate the participants to groups. Counterbalancing is often used with the repeated measures design in order to balance out any order effects and prevent them from disrupting the findings.

Good practice in nonexperimental designs

- We can distinguish between participant observation and nonparticipant observation research. Participant research involves the three stages of entering the field, being in the field, and leaving the field. Nonparticipant observation research involves devising precise, comprehensive, and usable behavioral categories. The sampling of behavior can be continuous, based on time intervals, based on time points, or random. Survey studies can use various designs: one-shot; before–after; two-groups controlled comparison; two groups before–after. When questionnaires or attitude scales are constructed, the items need to be short, unambiguous, and unbiased, and the tests need to be reliable and valid. Correlational studies should be based on an underlying theory, the variables should be carefully measured, and the scores on both variables should vary considerably from individual to individual.

Problems with experimental research

- Most research involves interactions between the experimenter and the participants. This can introduce various systematic biases, which can be divided into experimenter effects and demand characteristics. Experimenter effects include experimenter expectancy, experimenter misrecording, and experimenter fudging. Demand characteristics involve the participants responding on the basis of their beliefs about the experimental hypothesis or hypotheses. In addition, the behavior of participants is sometimes influenced by evaluation apprehension.

General issues in investigations

- In some studies, the independent variable has an effect on behavior simply because the participants know they are being observed. This is known as participant reactivity

or the Hawthorne effect. It is a serious problem, because it can lead us to misinterpret our findings. It is important for a study to have internal validity, meaning that the findings are genuine and caused by the independent variable. External validity, which refers to the extent to which the findings of a study can be generalized, is also important. Issues of generalizability apply to populations, locations, measures, and times. Information about the generalizability of any particular findings can be obtained by means of a meta-analysis. The measures used in a study should possess good reliability or consistency. If they do not, then they are inadequate measures of the variables in question, and it will be hard to replicate or repeat any findings obtained.

Further Reading

- Coolican, H. (2004). *Research methods and statistics in psychology* (4th ed). London: Hodder & Stoughton. Most of the topics discussed in this chapter are dealt with in a clear fashion in this textbook.
- Dyer, C. (1995). *Beginning research in psychology*. Oxford, UK: Blackwell. The various forms of nonexperimental study are discussed in a very accessible way by this author.

chapter 25

Contents

Qualitative analysis of data	585
Interpretation of interviews, case studies, and observations	587
Content analysis	589
Quantitative analysis: Descriptive statistics	591
Data presentation	596
Statistical tests	598
Issues of experimental and ecological validity	610
Writing up a practical	612

Data analysis

25

The data obtained from a study may or may not be in numerical or quantitative form, that is, in the form of numbers. If they are not in numerical form, then we can still carry out qualitative analyses based on the experiences of the individual participants. If they are in numerical form, then we typically start by working out some descriptive statistics to summarize the pattern of findings. These descriptive statistics include measures of central tendency within a sample (e.g., mean) and measures of the spread of scores within a sample (e.g., range). Another useful way of summarizing the findings is by means of graphs and figures. Several such ways of summarizing the data are discussed later on in this chapter.

In any study, two things might be true: (1) there is a difference (the experimental hypothesis), or (2) there is no difference (the null hypothesis). Various statistical tests have been devised to permit a decision between the experimental and null hypotheses on the basis of the data. Decision making based on a statistical test is open to error, in that we can never be sure whether we have made the correct decision. However, certain standard procedures are generally followed, and these are discussed in this chapter.

Finally, there are important issues relating to the validity of the findings obtained from a study. One reason why the validity of the findings may be limited is that the study itself was not carried out in a properly controlled and scientific fashion. Another reason why the findings may be partially lacking in validity is that they cannot readily be applied to everyday life, a state of affairs that occurs most often with laboratory studies. Issues relating to these two kinds of validity are discussed towards the end of the chapter.

QUALITATIVE ANALYSIS OF DATA

There is an important distinction between quantitative research and qualitative research. In quantitative research, the information obtained from the participants is expressed in numerical form. Studies in which we record the number of items recalled, reaction times, or the number of aggressive acts are all examples of quantitative research. In qualitative research, on the other hand, the information obtained from participants is *not* expressed in numerical form. The emphasis is on the stated experiences of the participants and on the stated meanings they attach to themselves, to other people, and to their environment. Those carrying out qualitative research sometimes make use of direct quotations from their participants, arguing that such quotations are often very revealing.

There has been rapid growth in the use of qualitative methods since the mid-1980s. This is due in part to increased dissatisfaction with the quantitative or scientific approach that has dominated psychology for the past 100 years. Coolican (1994) discussed a quotation from Reason and Rowan (1981), which expresses that dissatisfaction very clearly:

There is too much measurement going on. Some things which are numerically precise are not true; and some things which are not numerical are true. Orthodox research produces results which are statistically significant but humanly insignificant; in human inquiry it is much better to be deeply interesting than accurately boring.

Many experimental psychologists would regard this statement as being clearly an exaggeration. “Orthodox research” with its use of the experimental method has



Why do people behave in this way when on vacation? What motivates them to risk their health in the sun?

transformed our understanding of attention, perception, learning, memory, reasoning, and so on. However, qualitative research is of clear usefulness within some areas of social psychology, and it can shed much light on the motivations and values of individuals. As a result, investigators using interviews, case studies, or observations often make use of qualitative data, although they do not always do so.

Investigators who collect qualitative data use several different kinds of analysis, and so only general indications of what can be done with such data will be presented here. However, there would be general agreement among such investigators with the following statement by Patton (1980; cited in Coolican, 1994):

The cardinal principle of qualitative analysis is that causal relationships and theoretical statements be clearly emergent from and grounded in the phenomena studied. The theory emerges from the data; it is not imposed on the data.

How do investigators use this principle? One important way is by considering fully the categories spontaneously used by the participants *before* the investigators develop their own categories. An investigator first of all gathers together all the information obtained from the participants. This stage is not always entirely straightforward. For example, if we simply transcribe tape recordings of what our participants have said, we may be losing valuable information. Details about which words are emphasized, where the speaker pauses, and when the speaker speeds up or slows down should also be recorded, so that we can understand fully what he or she is trying to communicate.

The investigator then arranges the items of information (e.g., statements) into various groups in a preliminary way. If a given item seems of relevance to several groups, then it is included in all of them. Frequently, the next step is to take account of the categories or groupings suggested by the participants themselves. The final step is for the investigator to form a set of categories based on the information obtained from the previous steps. However, the investigator is likely to change some of the categories if additional information comes to light.

Qualitative investigators are not only interested in the number of items or statements falling into each category. Their major concern is usually in the variety of meanings, attitudes, and interpretations found within each category. For example, an investigator might study attitudes towards studying psychology by carrying out interviews with psychology students. One of the categories into which their statements were then placed might be “negative attitudes towards statistics.” A consideration of the various statements in this category might reveal numerous reasons why psychology students dislike statistics!

When qualitative researchers report their findings, they will often include some raw data (e.g., direct quotations from participants) as well as analyses of the data based on categories. In addition, they often indicate how their hypotheses changed during the course of the investigation.

EVALUATION

Qualitative analysis is often less influenced than is quantitative analysis by the biases and theoretical assumptions of the investigator. In addition, it offers the prospect of understanding the participants in a study as rounded individuals in a social context. This contrasts with quantitative analysis, in which the focus is often on rather narrow aspects of behavior.

The greatest limitation of the qualitative approach is that the findings that are reported tend to be unreliable and hard to replicate. Why is this so? The qualitative approach is subjective and impressionistic, and so the ways in which the information is categorized and then interpreted often differ considerably from one investigator to another.

There are various ways in which qualitative researchers try to show that their findings are reliable (Coolican, 1994). Probably the most satisfactory approach is to see whether the findings obtained from a qualitative analysis can be replicated. This can be done by comparing the findings from an interview study with those from an observational study.

Alternatively, two different qualitative researchers can conduct independent analyses of the same qualitative data, and then compare their findings.

Qualitative researchers argue that the fact that they typically go through the “research cycle” more than once helps to increase reliability. Thus, for example, the initial assumptions and categories of the researcher are checked against the data, and may then be changed. After that, the new assumptions and categories are checked against the data. Repeating the research cycle is of value in some ways, but it does not ensure that the findings will have high reliability.

INTERPRETATION OF INTERVIEWS, CASE STUDIES, AND OBSERVATIONS

Qualitative analyses as discussed in the previous section are carried out in several different kinds of studies. They are especially common in interviews, case studies, and observational studies, although quantitative analyses have often been used in all three types of studies. Some of the advantages and limitations of these types of studies are discussed in Chapter 23. What we do in this section is consider the interpretation of interviews, case studies, and observations.

INTERVIEWS

As discussed in Chapter 23, interviews vary considerably in terms of their degree of structure. In general terms, unstructured interviews (e.g., nondirective or informal) lend themselves to qualitative analyses, whereas structured interviews lend themselves to quantitative analysis. As Coolican (1994) pointed out, there are various skills that interviewers need in order to obtain valuable data. These skills involve establishing a good understanding with the person being interviewed, adopting a nonjudgmental approach, and developing effective listening skills.

Cardwell, Clark, and Meldrum (1996) illustrated the value of the interview approach by discussing the work of Reicher and Potter (1985) on a riot in the St. Paul’s area of Bristol in the UK in April 1980. Many of the media reports on the riot were based on the assumption that those involved in the riot were behaving in a primitive and excessively emotional way. Unstructured interviews with many of those involved indicated that in fact they had good reasons for their actions. They argued that they were defending their area against the police, and they experienced strong feelings of solidarity and community spirit. This interpretation was supported by the fact that very little of the damage affected private homes in the area.



Reicher and Potter argued that the St. Paul’s crowd saw themselves as a legitimate presence and the police as an illegitimate presence. Each group attached a different meaning to their actions. Does that make interpretation of discourse problematic?

Evaluation

There are various problems involved in interpreting interview information.

First, there is the problem of **social desirability bias**. Most people want to present themselves in the best possible light, so they may provide socially desirable rather than honest answers to personal questions. This problem can be handled by the interviewer asking additional questions to establish the truth.

Second, the data obtained from an interviewer may reveal more about the social interaction processes between the interviewer and the person being interviewed (the interviewee) than about the interviewee’s thought processes and attitudes.

Third, account needs to be taken of the **self-fulfilling prophecy**. This is the tendency for someone’s expectations about another person to lead to the fulfillment of those expectations. For example, suppose that a therapist expects his or her patient to behave very anxiously. This expectation may cause the therapist to treat the patient in such a way that the patient starts to behave in the expected fashion.

Key Terms

Social desirability bias: the tendency to provide socially desirable rather than honest answers on questionnaires and in interviews.

Self-fulfilling prophecy: the tendency for someone’s expectations about another person to lead to the fulfillment of those expectations.

CASE STUDIES

Case studies (intensive investigations of individuals) come in all shapes and sizes. Probably the best-known case studies are those of Freud and others in the field of clinical psychology. However, detailed case studies have also been carried out in personality research and in studies of cognitive functioning in brain-damaged patients.

One way in which case studies have been used to study personality involves an approach known as **psychobiography**. This was defined by McAdams (1988, p. 2) as “the systematic use of psychological (especially personality) theory to transform a life into a coherent and illuminating story.” A key feature of psychobiography is identification of the most important events in an individual’s account of his or her own life story. How can this be done? According to McAdams (1988, pp. 12–13), we should look for

clues about primacy (what comes first in a story), uniqueness (what stands out in the story), omission (what seems to be missing from the story), distortion and isolation (what doesn’t follow logically in the story), and incompleteness (when the story fails to end in a satisfying way).

Weiskrantz (1986) reported a very different kind of case study. He studied DB, who had had an operation designed to reduce the number of severe migraines from which he suffered. As a result of this operation, DB exhibited what is known as “blindsight.” He was able to tell whether a visual stimulus had been presented, and he could point at it, even though he had no conscious awareness of having seen it. These findings are important, because they suggest that many perceptual processes can occur in the absence of conscious awareness.

Case Study: The Effects of Extreme Deprivation

Freud and Dann (1951) studied six preschool children who had lost their parents during the Second World War. It is not known how long each child had spent with their parents before being taken to Nazi concentration camp nurseries. The children remained together, despite moving camp several times, and appeared to have received only the most basic forms of care and attention. In the absence of a caring adult, they had formed close and loving bonds with each other. These strong bonds provided a protective and stable influence in their lives.

The children were rescued at the end of the war and brought to England for medical and psychological treatment. Their mental and physical development had been restricted, so that they had very poor speech skills. They feared adults and clung to each other for reassurance. Gradually they began to form bonds with the adults who cared for them, and their social and language skills improved.

Despite all the problems they had experienced, the children did not show the levels of extreme disturbance that were once expected when there is a complete lack of “mothering” (Bowlby, 1951). Freud and Dann’s study highlights the fundamental importance of having someone to bond with, even if it is not the mother, as well as the reversibility of the effects of extreme deprivation.

Case studies are often seen as rather unscientific and unreliable. The sample is not representative of the wider population, the study cannot be repeated, and interpretation of the findings is very subjective. However, case studies can be of great interest because they highlight unique and unexpected behavior, and can stimulate research that may contradict established theories such as Bowlby’s. Freud and Dann’s work offers insights into human experience that would otherwise be impossible to gain: ethical considerations prevent the deliberate separation of children and parents in order to study the effects of deprivation.

Evaluation

We need to be very careful when interpreting the evidence from a case study. The greatest danger is that very general conclusions may be drawn on the basis of a single atypical individual. For this reason, it is important to have supporting evidence from other sources before drawing such conclusions.

It is often hard to interpret the evidence from case studies. For example, Freud claimed that the various case studies he reported served to show the validity of his theoretical ideas. However, such evidence is suspect, because there was a real chance of

Key Term

Psychobiography: the study of individual personality by applying psychological theory to the key events in a person’s life.

contamination in the data Freud obtained from his patients. What any patient said to Freud may have been influenced by what Freud had said to him or her previously, and Freud may have used his theoretical views to interpret what the patient said in ways that distorted it.

How, then, should the findings from a case study be interpreted? Probably the greatest value of a case study is that it can suggest hypotheses that can then be tested under more controlled conditions with larger numbers of participants. In other words, case studies usually provide suggestive rather than definitive evidence. In addition, case studies can indicate that there are limitations in current theories. The discovery of blindsight in DB suggested that visual perception depends much less on conscious awareness than was thought to be the case by most theorists.

OBSERVATIONS

As discussed in Chapter 23, there are numerous kinds of observational studies, and the data obtained may be either quantitative or qualitative. We consider issues relating to interpreting the data from observational studies by focusing on a concrete example. Jourard (1966) watched pairs of people talking in cafes, and noted down the number of times one person touched another at one table during 1 hour. In San Juan, the capital of Puerto Rico, the total number of touches was 180. In contrast, the total in Paris was 110, and in London it was 0. One problem with interpreting these data is that the kinds of people who go to cafes in San Juan, Paris, and London may be quite different. It is also entirely possible that those who spend much of their time in cafes are not representative of the general population. These issues of representativeness apply to many observational studies.

Evaluation

Jourard's (1966) findings do not really tell us why there is (or was, in 1966) much more touching in San Juan than in London. It is possible that Londoners are simply less friendly and open, but there are several other possibilities (e.g., Londoners are more likely to go to cafes with business colleagues). The general issue here is that it is often very hard to interpret or make sense of the data obtained from observational studies, because we can only speculate on the reasons why the participants are behaving in the ways that we observe.

Another issue was raised by Coolican (1994) in his discussion of the work of Whyte (1943). Whyte joined an Italian street gang in Chicago, and became a participant observer. The problem he encountered in interpreting his observations was that his presence in the gang influenced their behavior. A member of the gang expressed this point as follows: "You've slowed me down plenty since you've been down here. Now, when I do something, I have to think what Bill Whyte would want me to know about it and how I can explain it."

CONTENT ANALYSIS

Content analysis is used when originally qualitative information is reduced to numerical terms. **Content analysis** started off as a method for analyzing messages in the media, including articles published in newspapers, speeches made by politicians on radio and television, various forms of propaganda, and health records. More recently, the method of content analysis has been applied more widely to almost any form of communication. As Coolican (1994, p. 108) pointed out:

The communications concerned were originally those already published, but some researchers conduct content analysis on materials which they ask people to produce, such as essays, answers to interview questions, diaries, and verbal protocols [detailed records].

One of the types of communication that has often been studied by content analysis is television advertising. For example, McArthur and Resko (1975) carried out a content

Key Term

Content analysis: a method involving the detailed study of, for example, the output of the media, speeches, and literature.



Content analysis of advertising can tell us a great deal about society's attitudes to men and women.

analysis of American television commercials. They found that 70% of the men in these commercials were shown as experts who knew a lot about the products being sold. In contrast, 86% of the women in the commercials were shown only as product users. There was another interesting gender difference: Men who used the products were typically promised improved social and career prospects, whereas women were promised that their family would like them more.

More recent studies of American television commercials (e.g., Brett & Cantor, 1988) indicate that the differences in the ways in which men and women are presented have been reduced. However, it remains the case that the men are far more likely than women to be presented as the product expert.

The first stage in content analysis is that of sampling, or deciding what to select from what may be an enormous amount of material. For example, when Cumberbatch (1990) carried out a study on over 500 advertisements shown on British television, there were two television channels showing advertisements. Between them, these two channels were broadcasting for about 15,000 hours a year, and showing over 250,000 advertisements. Accordingly, Cumberbatch decided to select only a sample of advertisements taken from prime-time television over a 2-week period.

The issue of sampling is an important one. For example, television advertisers target their advertisements at particular sections of the population, and so arrange for the advertisements to be shown when the relevant groups are most likely to be watching television. As a result, advertisements for beer are more likely to be shown during a soccer match than during a program about fashion. By focusing on prime-time television, Cumberbatch (1990) tried to ensure that he was studying advertisements designed to have general appeal.

The other key ingredient in content analysis is the construction of the **coding units** into which the information is to be categorized. In order to form appropriate coding units, the researcher needs to have considerable knowledge of the kinds of material to be used in the content analysis. He or she also needs to have one or more clear hypotheses, because the selection of coding units must be such as to permit these hypotheses to be tested effectively.

The coding can take many forms. The categories used can be very specific (e.g., use of a given word) or general (e.g., theme of the communication). Instead of using categories, the coders may be asked to provide *ratings*. For example, the apparent expertise of those appearing in television advertisements might be rated on a 7-point scale. Another form of coding involves *ranking* items, or putting them in order. For example, the statements of politicians could be ranked in terms of the extent to which they agreed with the facts.

Key Term

Coding units: the categories into which observations are placed prior to analysis.

EVALUATION

One of the greatest strengths of content analysis is that it provides a way of extracting information from a wealth of real-world settings. The media influence the ways we think and feel about issues, and so it is important to analyze media communications in detail. Content analysis can reveal issues of concern. For example, Cumberbatch (1990) found

in his study of advertisements on British television that only about 25% of the women appearing in these advertisements seemed to be over 30 years old, compared to about 75% of the men. On the face of it, this would seem to reflect a sexist bias.

The greatest limitation of content analysis is that it is often very hard to interpret the findings. Consider, for example, the difference in the ages of men and women appearing in advertisements found by Cumberbatch (1990). One interpretation is that this difference occurred because most television viewers prefer to see older men and younger women in advertisements. However, it is also possible that those making the advertisements thought mistakenly that this is what the viewers wanted to see.

Gender and advertising

Cumberbatch (1990) found that men outnumbered women in advertisements by 2 : 1. In addition, 75% of the men in ads were aged over 30, whereas 75% of women in ads were aged under 30. Male voices were used where the information in the soundtrack concerned technical expertise, whereas women's voices were used in sexy and sensuous ways. What does this say about the way we view men and women in society? Comparing the results of studies such as Cumberbatch's with earlier ones (e.g., McArthur & Resko, 1975) can begin to provide answers to questions such as this.

Food Diary – Week 1

Time	What eaten	B	V	L	Antecedents & Consequences
8.00	All-brown				A: Still full from yesterday. C: Must make an effort not to binge today.
12.00	1 apple				A: Hungry. C: Still hungry, mustn't eat more in case it starts me off on a binge.
3.00	1 lb grapes, 2 choc. bars		!		A: Had phone call from John, he will be home late. C: Disgusted with myself. I am the most hopeless person in the world.
6.00	peanuts + choc, picked from shopping	!!			A: No food in flat. Had to go shopping. Couldn't stop myself putting loads of sweets in the trolley. Ate loads of stuff in the car. Had to go on eating once at home.
7.00	2 portions of corn, 3 choc. bars	!!		!!	C: Very angry with myself. I feel so lonely. Totally exhausted, went to bed early.

B = Binge, V = Vomited, L = Laxatives

Food Diary – Week 4

Time	What eaten	B	V	L	Antecedents & Consequences
8.00	Cottage cheese, 2 sl. toast with honey				Enjoyed this.
11.00	apple				
12.30	baked potato, tuna fish				Eaten in the canteen at work. Tina said "You haven't been here for ages". Could have run away, felt everybody was looking at me.
3.00	yaourt, crunch bar				
6.00	1 sl. toast				
7.00	fish + vegetables, 1 portion ice cream				Had not planned dessert. John suggested ice cream. My initial response was to say no, but I know I would then finish the packet off whilst washing up. So I had a portion and enjoyed it sitting with John. John put it away and made coffee, which we drank relaxing on the sofa. Washing up left.

Diary studies are often used in clinical psychology, such as in this example from the diary of a bulimia sufferer. Diaries may be used to record actions, thoughts, and feelings, but may not be totally accurate, particularly if the diarist is embarrassed to reveal the truth about himself or herself. Food diaries reproduced from Schmidt and Treasure (1993), with permission. Copyright © 1993 Psychology Press.

There are other possible interpretations, but the available data do not allow us to discriminate among them.

There are also problems of interpretation with other communications such as personal diaries or essays. Diaries or essays may contain accurate accounts of what an individual does, thinks, and feels. On the other hand, individuals may provide deliberately distorted accounts in order to protect their self-esteem, to make it appear that their lives are more exciting than is actually the case, and so on.

Another problem is that the selection and scoring of coding units can be rather subjective. The coding categories that are used need to accurately reflect the content of the communication, and each of the categories must be defined as precisely as possible.

QUANTITATIVE ANALYSIS: DESCRIPTIVE STATISTICS

Suppose that we have carried out an experiment on the effects of noise on learning with three groups of nine participants each. One group was exposed to very loud noise, another group to moderately loud noise, and the third group was not exposed to noise at all. What they had learned from a book chapter was assessed by giving them a set of questions, producing a score between 0 and 20.

What is to be done with the raw scores? There are two key types of measures that can be taken whenever we have a set of scores from participants in a given condition. First, there are measures of central tendency, which provide some indication of the size of average or typical scores. Second, there are measures of dispersion, which indicate the extent to which the scores cluster around the average or are spread out. Various measures of central tendency and of dispersion are considered next.

MEASURES OF CENTRAL TENDENCY

Measures of central tendency describe how the data cluster together around a central point. There are three main measures of central tendency: the mean; the median; and the mode.