VOL. 63, NO. 5 SEPTEMBER, 1956

THE PSYCHOLOGICAL REVIEW

STRUCTURAL BALANCE: A GENERALIZATION OF HEIDER'S THEORY [1]

DORWIN CARTWRIGHT AND FRANK HARARY
Research Center for Group Dynamics, University of Michigan

A persistent problem of psychology has been how to deal conceptually with patterns of interdependent properties. This problem has been central, of course, in the theoretical treatment by Gestalt psychologists of phenomenal or neural *configurations* or *fields* (**12**, **13**, **15**). It has also been of concern to social psychologists and sociologists who attempt to employ concepts referring to social *systems* (**18**).

Heider (**19**), reflecting the general field-theoretical approach, has considered certain aspects of cognitive fields which contain perceived people and impersonal objects or events. His analysis focuses upon what he calls the *P-O-X* unit of a cognitive field, consisting of *P* (one person), *O* (another person), and *X* (an impersonal entity). Each relation among the parts of the unit is conceived as interdependent with each other relation. Thus, for example, if *P* has a relation of affection for *O* and if *O* is seen as responsible for *X*, then there will be a tendency for *P* to like or approve of *X*. If the nature of *X* is such that it would "normally" be evaluated as bad, the whole *P-O-X* unit is placed in a state of imbalance, and pressures will arise to change it toward a state of balance. These pressures may work to change the relation of affection between *P* and *O*, the relation of responsibility between *O* and *X*, or the relation of evaluation between *P* and *X*.

The purpose of this paper is to present and develop the consequences of a formal definition of balance which is consistent with Heider's conception and which may be employed in a more general treatment of empirical configurations. The definition is stated in terms of the mathematical theory of linear graphs (**8**, **14**) and makes use of a distinction between a given relation and its opposite relation. Some of the ramifications of this definition are then examined by means of theorems derivable from the definition and from graph theory.

HEIDER'S CONCEPTION OF BALANCE

In developing his analysis of balanced cognitive units, Heider distinguishes between two major *types* of relations. The first concerns attitudes, or the relation of liking or evaluating. It is represented symbolically as **L** when positive and as ~**L** when negative. Thus, *P***L***O* means *P* likes, loves, values, or approves *O*, and *P*~**L***O* means *P* dislikes, negatively

[1] This paper was prepared as part of a project sponsored in the Research Center for Group Dynamics by the Rockefeller Foundation.

values, or disapproves *O*. The second type of relation refers to cognitive unit formation, that is, to such specific relations as similarity, possession, causality, proximity, or belonging. It is written as **U** or ∼**U**. Thus, according to Heider, *P***U***X* means that *P* owns, made, is close to, or is associated with *X*, and *P*∼**U***X* means that *P* does not own, did not make, or is not associated with *X*.

A *balanced state* is then defined in terms of certain combinations of these relations. The definition is stated separately for two and for three entities.

> In the case of two entities, a balanced state exists if the relation between them is positive (or negative) in all respects, i.e., in regard to all meanings of **L** and **U** . . . . In the case of three entities, a balanced state exists if all three relations are positive in all respects, or if two are negative and one positive (9, p. 110).

These are examples of balanced states: *P* likes something he made (*P***U***X*, *P***L***X*); *P* likes what his friend likes (*P***L***O*, *O***L***X*, *P***L***X*); *P* dislikes what his friend dislikes (*P***L***O*, *O*∼**L***X*, *P*∼**L***X*); *P* likes what his enemy dislikes (*P*∼**L***O*, *O*∼**L***X*, *P***L***X*); and *P*'s son likes what *P* likes (*P***U***O*, *P***L***X*, *O***L***X*).

Heider's basic hypothesis asserts that there is a tendency for cognitive units to achieve a balanced state. Pressures toward balance may produce various effects.

> If no balanced state exists, then forces towards this state will arise. Either the dynamic characters will change, or the unit relations will be changed through action or through cognitive reorganization. If a change is not possible, the state of imbalance will produce tension (9, pp. 107–109).

The theory, stated here in sketchy outline, has been elaborated by Heider so as to treat a fuller richness of cognitive experience than would be suggested by our brief description. It has been used, too, by a number of others as a point of departure for further theoretical and empirical work. We shall summarize briefly some of the major results of this work.

Horowitz, Lyons, and Perlmutter (**10**) attempted to demonstrate tendencies toward balance in an experiment employing members of a discussion group as subjects. At the end of a discussion period each subject was asked to indicate his evaluation of an event (*P***L***X* or *P*∼**L***X*) which had occurred during the course of the discussion. The event selected for evaluation was one which would be clearly seen as having been produced by a single person (*O***U***X*). The liking relation between each *P* and *O* (*P***L***O* or *P*∼**L***O*) had been determined by a sociometric questionnaire administered before the meeting. Would *P*'s evaluation of the event be such as to produce a balanced *P-O-X* unit? If so, *P*'s evaluation of *O* and *X* should be of the same sign. The experimental data tend to support the hypothesis that a *P-O-X* unit tends toward a balanced state.[2]

The social situation of a discussion group can be better analyzed, according to Horowitz, Lyons, and Perlmutter, by considering a somewhat more complex cognitive unit. The evaluation of *X* made by *P*, they argue, will be determined not only by *P*'s evaluation of *O* but also by his perception of the evaluation of *X* given by others (*Q*s) in the group. The basic unit of such a social situation, then, consists of the subject, a

[2] One of the attractive features of this study is that it was conducted in a natural "field" setting, thus avoiding the dangers of artificiality. At the same time the setting placed certain restrictions on the possibility of manipulation and control of the variables. The data show a clear tendency for *P* to place a higher evaluation on *X*s produced by more attractive *O*s. It is not clearly demonstrated that *P* likes *X*s produced by liked *O*s and dislikes *X*s produced by disliked *O*s.

person who is responsible for the event, and another person who will be seen by the subject as supporting or rejecting the event. This is called a *P-O-Q-X* unit. The additional data needed to describe these relations were obtained from the sociometric questionnaire which indicated *P*'s evaluation of *Q* (*P*L*Q* or *P*$\sim$L*Q*), and from a question designed to reveal *P*'s perception of *Q*'s support or rejection of *X*, treated by the authors as a unit relation (*Q*U*X* or *Q*$\sim$U*X*).[3]

Although these authors indicate the possibility of treating the *P-O-Q-X* unit in terms of balance, they do not develop a formal definition of a balanced configuration consisting of four elements. They seem to imply that the *P-O-Q-X* unit will be balanced if the *P-O-X* and the *P-Q-X* units are both balanced. They do not consider the relation between *Q* and *O*, nor the logically possible components of which it could be a part. Their analysis is concerned primarily with the two triangles (*P-O-X* and *P-Q-X*), which are interdependent, since both contain the relation of *P*'s liking of *X*. We noted above that the data tend to support the hypothesis that the *P-O-X* unit will tend toward balance. The data even more strongly support the hypothesis when applied to the *P-Q-X* unit; *P*'s evaluation of *X* and his perception of *Q*'s attitude toward *X* tend to agree when *P* likes *Q*, and to disagree when *P* dislikes *Q*. It should be noted, however, that there was also a clear tendency for *P* to see *Q*'s evaluation of *X* as agreeing with his own whether or not he likes *Q*.

In a rather different approach to the question of balanced *P-O-X* units, Jordan (**11**) presented subjects with 64 different hypothetical situations in which the **L** and **U** relations between each pair of elements was systematically varied. The subject was asked to place himself in each situation by taking the part of *P*, and to indicate on a scale the degree of pleasantness or unpleasantness he experienced. Unpleasantness was assumed to reflect the postulated tension produced by imbalanced units. Jordan's data tend to support Heider's hypothesis that imbalanced units produce a state of tension, but he too found that additional factors need to be considered. He discovered, for example, that negative relations were experienced as unpleasant even when contained in balanced units. This unpleasantness was particularly acute when *P* was a part of the negative relation. Jordan's study permits a detailed analysis of these additional influences, which we shall not consider here.

Newcomb (**17**), in his recent theory of interpersonal communication, has employed concepts rather similar to those of Heider. He conceives of the simplest communicative act as one in which one person *A* gives information to another person *B* about something *X*. The similarity of this *A-B-X* model to Heider's *P-O-X* unit, together with its applicability to objective interpersonal relations (rather than only to the cognitive structure of a single person), may be seen in the following quotations from Newcomb:

> *A-B-X* is . . . regarded as constituting a system. That is, certain definable relationships between *A* and *B*, between *A* and *X*, and between *B* and *X* are all viewed as interdependent. . . . For some purposes the system may be regarded as a phenomenal one within the life space of *A* or *B*, for other purposes as an "objective" system including all the possible relationships as inferred from observations of *A*'s and *B*'s behavior (**17**, p. 393).

[3] Whether this relation should be treated as **U** or **L** is subject to debate. For testing Heider's theory of balance, however, the issue is irrelevant, since he holds that the two relations are interchangeable in defining balance.

Newcomb then develops the concept of "strain toward symmetry," which appears to be a special instance of Heider's more general notion of "tendency toward balance." "Strain toward symmetry" is reflected in several manifestations of a tendency for *A* and *B* to have attitudes of the same sign toward a common *X*. Communication is the most common and usually the most effective manifestation of this tendency.

By use of this conception Newcomb reinterprets several studies (**1**, **4**, **5**, **16**, **20**) which have investigated the interrelations among interpersonal attraction, tendencies to communicate, pressures to uniformity of opinion among members of a group, and tendencies to reject deviates. The essential hypothesis in this analysis is stated thus:

> If *A* is free either to continue or not to continue his association with *B*, one or the other of two eventual outcomes is likely: (*a*) he achieves an equilibrium characterized by relatively great attraction toward *B* and by relatively high perceived symmetry, and the association is continued; or (*b*) he achieves an equilibrium characterized by relatively little attraction toward *B* and by relatively low perceived symmetry, and the association is discontinued (**17**, p. 402).

Newcomb's outcome *a* is clearly a balanced state as defined by Heider. Outcome *b* cannot be unambiguously translated into Heider's terms. If by "relatively little attraction toward *B*" is meant a negative **L** relation between *A* and *B*, then this outcome would also seem to be balanced. Newcomb's "continuation or discontinuation of the association between *A* and *B*" appear to correspond to Heider's **U** and ~**U** relations.

STATEMENT OF THE PROBLEM

This work indicates that the tendency toward balance is a significant determinant of cognitive organization, and that it may also be important in interpersonal relations. The concept of balance, however, has been defined so as to apply to a rather limited range of situations, and it has contained certain ambiguities. We note five specific problems.

1. *Unsymmetric relations*. Should all relations be conceived as symmetric? The answer is clearly that they should not; it is possible for *P* to like *O* while *O* dislikes *P*. In fact, Tagiuri, Blake, and Bruner (**21**) have intensively studied dyadic relations to discover conditions producing symmetric relations of actual and perceived liking. Theoretical discussions of balance have sometimes recognized this possibility—Heider, for example, states that unsymmetric liking is unbalanced—but there has been no general definition of balance which covers unsymmetric relations. The empirical studies of balance have assumed that the relations are symmetric.

2. *Units containing more than three entities*. Nearly all theorizing about balance has referred to units of three entities. While Horowitz, Lyons, and Perlmutter studied units with four entities, they did not *define* balance for such cases. It would seem desirable to be able to speak of the balance of even larger units.

3. *Negative relations*. Is the negative relation the *complement* of the relation or its *opposite*? All of the discussions of balance seem to equate these, but they seem to us to be quite different, for the complement of a relation is expressed by adding the word "not" while the opposite is indicated by the prefix "dis" or its equivalent. Thus, the complement of "liking" is "not liking"; the opposite of "liking" is "disliking." In general, it appears that ~**L** has been taken to mean "dislike" (the opposite relation) while

~U has been used to indicate "not associated with" (the complementary relation). Thus, for example, Jordan says: "Specifically, '+L' symbolizes a positive attitude, '−L' symbolizes a negative attitude, '+U' symbolizes the existence of unit formation, and '−U' symbolizes the lack of unit formation" (**11**, p. 274).

4. *Relations of different types.* Heider has made a distinction between two types of relations—one based upon liking and one upon unit formation. The various papers following up Heider's work have continued to use this distinction. And it seems reasonable to assume that still other types of relations might be designated. How can a definition of balance take into account relations of different types? Heider has suggested some of the ways in which liking and unit relations may be combined, but a general formulation has yet to be developed.

5. *Cognitive fields and social systems.* Heider's intention is to describe balance of cognitive units in which the entities and relations enter as experienced by a single individual. Newcomb attempts to treat social systems which may be described objectively. In principle, it should be possible also to study the balance of sociometric structures, communication networks, patterns of power, and other aspects of social systems.

We shall attempt to define balance so as to overcome these limitations. Specifically, the definition should (*a*) encompass unsymmetric relations, (*b*) hold for units consisting of any finite number of entities, (*c*) preserve the distinction between the *complement* and the *opposite* of a relation, (*d*) apply to relations of different types, and (*e*) serve to characterize cognitive units, social systems, or any configuration where both a relation and its opposite must be specified.

THE CONCEPTS OF GRAPH, DIGRAPH, AND SIGNED GRAPH

Our approach to this problem has two primary antecedents: (*a*) Lewin's treatment (**15**) of the concepts of whole, differentiation, and unity, together with Bavelas' extension (**2**) of this work to group structure; and (*b*) the mathematical theory of linear graphs.

Many of the graph-theoretic definitions given in this section are contained in the classical reference on graph theory, König (**14**), as well as in Harary and Norman (**8**). We shall discuss, however, those concepts which lead up to the theory of balance.

A *linear graph*, or briefly a *graph*, consists of a finite collection of *points* [4] $A, B, C, \cdots$ together with a prescribed subset of the set of all unordered pairs of distinct points. Each of these unordered pairs, AB, is a *line* of the graph. (From the viewpoint of the theory of binary relations,[5] a graph corresponds to an irreflexive [6] symmetric relation on points A, B, C, $\cdots$. Alternatively a graph may be represented as a matrix.[7])

Figure 1 depicts a graph of four points and four lines. The points might represent people, and the lines some relationship such as mutual liking. With this interpretation, Fig. 1 indicates that mutual liking exists between those pairs of people A, B, C, and D joined by lines. Thus D is in the relation with all other persons, while C is in the relation only with D.

[4] Points are often called "vertices" by mathematicians and "nodes" by electrical engineers.

[5] This is the approach used by Heider.

[6] A relation is irreflexive if it contains no ordered pairs of the form (a, a), i.e., if no element is in this relation to itself.

[7] This treatment is discussed in Festinger (**3**). The logical equivalence of relations, graphs, and matrices is taken up in Harary and Norman (**8**).

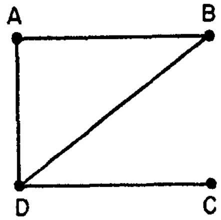

FIG. 1. A linear graph of four points and four lines. The presence of line AB indicates the existence of a specified symmetric relationship between the two entities A and B.

Figure 1 could be used, of course, to represent many other kinds of relationships between many other kinds of entities.

It is apparent from this definition of graph that relations are treated in an all-or-none manner, i.e., either a relation exists between a given pair of points or it does not. Obviously, however, many relationships of interest to psychologists (liking, for example) exist in varying degrees. This fact means that our present use of graph theory can treat only the structural, and not the numerical, aspects of relations. While our treatment is thereby an incomplete representation of the strength of relations, we believe that conceptualization of the structural properties of relations is a necessary first step toward a more adequate treatment of the more complex situations. Such an elaboration, however, goes beyond the scope of this paper.

A *directed graph*, or a *digraph*, consists of a finite collection of points together with a prescribed subset of the set of all ordered pairs of distinct points. Each of these ordered pairs $\overrightarrow{AB}$ is called a *line* of the digraph. Note that the only difference between the definitions of graph and digraph is that the lines of a graph are unordered pairs of points while the lines of a digraph are ordered pairs of points. An *ordered pair* of points is distinguished from an unordered pair by designating one of the points as the first point and the other as the second. Thus, for example, the fact that a message can go *from A to B* is represented by the ordered pair (A, B), or equivalently, by the line $\overrightarrow{AB}$, as in Fig. 2. Similarly, the fact that A and D choose each other is represented by the two directed lines $\overrightarrow{AD}$ and $\overrightarrow{DA}$.

A *signed graph*, or briefly an *s-graph*, is obtained from a graph when one regards some of the lines as positive and the remaining lines as negative. Considered as a geometric representation of binary relations, an s-graph serves to depict situations or structures in which both a relation and its opposite may occur, e.g., like and dislike. Figure 3 depicts an s-graph, employing the convention that solid lines are positive and dashed lines negative; thus A and B are represented as liking each other while A and C dislike each other.

Combining the concepts of digraph and s-graph, we obtain that of an s-digraph. A *signed digraph*, or an *s-digraph*, is obtained from a digraph by taking some of its lines as positive and the rest as negative.

A *graph of type 2* (8), introduced to depict structures in which two differ-

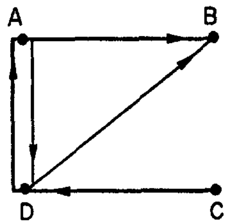

FIG. 2. A directed graph of four points and five directed lines. An $\overrightarrow{AB}$ line indicates the existence of a specified ordered relationship involving the two entities A and B. Thus, for example, if A and B are two people, the $\overrightarrow{AB}$ line might indicate that a message can go from A to B or that A chooses B.

ent relations defined on the same set of elements occur, is obtained from a graph by regarding its lines as being of two different colors (say), and by permitting the same pair of points to be joined by two lines if these lines have different colors. A *graph of type* τ, $\tau = 1, 2, 3, \cdots$, is defined similarly. In an s-graph or s-digraph of type 2, there may occur lines of two different types in which a line of either color may be positive or negative. An example of an s-graph of type 2 might be one depicting for the same *P-O-X* unit both **U** and **L** relations among the entities, where the sign of these relations is indicated.

A *path* is a collection of lines of a graph of the form $AB, BC, \cdots, DE$, where the points $A, B, C, \cdots, D, E$, are distinct. A *cycle* consists of the above path together with the line EA. The *length* of a cycle (or path) is the number of lines in it; an *n-cycle* is a cycle of length n. Analogously to graphs, a *path of a digraph* consists of directed lines of the form $\overrightarrow{AB}, \overrightarrow{BC}, \cdots, \overrightarrow{DE}$, where the points are distinct. A *cycle* consists of this path together with the line $\overrightarrow{EA}$. In the later discussion of balance of an s-digraph we shall use the concept of a semicycle. A *semicycle* is a collection of lines obtained by taking exactly one from each pair $\overrightarrow{AB}$ or $\overrightarrow{BA}$, $\overrightarrow{BC}$ or $\overrightarrow{CB}, \cdots, \overrightarrow{DE}$ or $\overrightarrow{ED}$, and $\overrightarrow{EA}$ or $\overrightarrow{AE}$. We illustrate semicycles with the digraph of Fig. 2. There are three semicycles in this digraph: $\overrightarrow{AD}, \overrightarrow{DA}$; $\overrightarrow{AD}, \overrightarrow{DB}, \overleftarrow{BA}$; and $\overleftarrow{AD}, \overleftarrow{DB}, \overleftarrow{BA}$. The last two of these semicycles are not cycles. Note that every cycle is a semicycle, and a semicycle of length 2 is necessarily a cycle.

BALANCE

With these concepts of graphs, digraphs, and signed graphs we may now develop a rigorous generalization of Heider's concept of balance.

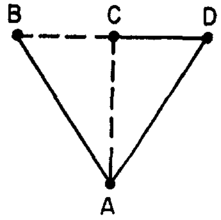

FIG. 3. A signed graph of four points and five lines. Solid lines have a positive sign and dashed lines a negative sign. If the points stand for people and the lines indicate the existence of a liking relationship, this s-graph shows that A and B have a relationship of liking, A and C have one of disliking, and B and D have a relationship of indifference (neither liking or disliking).

It should be evident that Heider's terms, *entity*, *relation*, and *sign of a relation* may be coordinated to the graphic terms, *point*, *directed line*, and *sign of a directed line*. Thus, for example, the assertion that P likes O (P**L**O) may be depicted as a directed line of positive sign $\overrightarrow{PO}$. It should also be clear that Heider's two different kinds of relations (**L** and **U**) may be treated as lines of different type. It follows that a graphic representation of a *P-O-X* unit having positive or negative **L** and **U** relations will be an s-digraph of type 2.

For simplicity of discussion we first consider the situation containing only symmetric relations of a single type (i.e., an s-graph of type 1). Figure 4 shows four such s-graphs. It will be noted that each of these s-graphs contains one cycle: AB, BC, CA. We now need to define the sign of a cycle. The *sign of a cycle* is the product of the signs of its lines. For convenience we denote the sign of a line by $+1$ or -1 when it is positive or negative. With this definition we see that the cycle, AB, BC, CA is positive in s-graph a $(+1\cdot+1\cdot+1)$, positive in s-graph b $(+1\cdot-1\cdot-1)$, negative

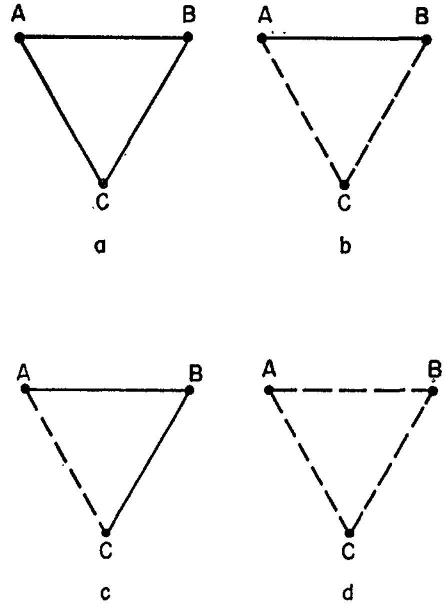

FIG. 4. Four s-graphs of three points and three lines each. Structure *a* and *b* are balanced, but *c* and *d* are not balanced.

in s-graph *c* $(+1\cdot+1\cdot-1)$, and negative in s-graph *d* $(-1\cdot-1\cdot-1)$. To generalize, a *cycle is positive* if it contains an even number of negative lines, and it is *negative* otherwise. Thus, in particular, a cycle containing only positive lines is positive, since the number of negative lines is zero, an even number.

In discussing the concept of balance, Heider states (see **9**, p. 110) that when there are three entities a balanced state exists if all three relations are positive or if two are negative and one positive. According to this definition, s-graphs *a* and *b* are balanced while s-graphs *c* and *d* are not (Fig. 4). We note that in the examples cited Heider's balanced state is depicted as an s-graph of three points whose cycle is positive.

In generalizing Heider's concept of balance, we propose to employ this characteristic of balanced states as a general criterion for balance of structures with any number of entities. Thus we define an *s-graph* (containing any number of points) as *balanced* if all of its cycles are positive.

Figure 5 illustrates this definition for four s-graphs containing four points. In each of these s-graphs there are seven cycles: *AB*, *BC*, *CA*; *AB*, *BD*, *DA*; *BC*, *CD*, *DB*; *AC*, *CD*, *DA*; *AB*, *BC*, *CD*, *DA*; *AB*, *BD*, *DC*, *CA*; and *BC*, *CA*, *AD*, *DB*. It will be seen that in s-graphs *a* and *b* all seven cycles are positive, and these s-graphs are therefore balanced. In s-graphs *c* and *d* the cycle, *AB*, *BC*, *CA*, is negative (as are several others), and these s-graphs are therefore not balanced. It is obvious that this definition of balance is applicable to structures containing any number of entities.[8]

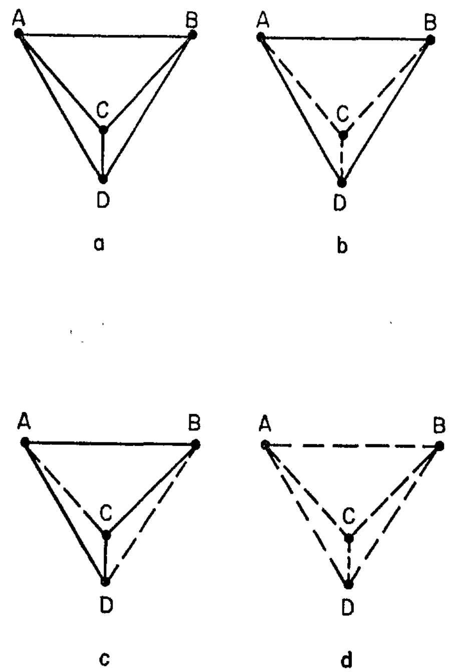

FIG. 5. Four s-graphs containing four points and six lines each. Structures *a* and *b* are balanced, but *c* and *d* are not balanced.

[8] If an s-graph contains no cycles, we say that it is "vacuously" balanced, since all (in this case, none) of its cycles are positive.

The extension of this definition of balance to s-digraphs containing any number of points is straightforward. Employing the same definition of *sign of a semicycle* for an s-digraph as for an ordinary s-graph, we similarly define an s-digraph as balanced if all of its semicycles are positive.

Consider now Heider's *P-O-X* unit, containing two persons *P* and *O* and an impersonal entity *X*, in which we are concerned only with liking relations. Figure 6 shows three of the possible 3-point s-digraphs which may represent such *P-O-X* units. A positive $\overrightarrow{PO}$ line means that *P* likes *O*, a negative $\overrightarrow{PO}$ line means that *P* dislikes *O*. We assume that a person can like or dislike an impersonal entity but that an impersonal entity can neither like nor dislike a person.[9] We also rule out of consideration here "ambivalence," where a person may simultaneously like and dislike another person or impersonal entity.

In each of these s-digraphs there are three semicycles: $\overrightarrow{PO}$, $\overrightarrow{OP}$; $\overrightarrow{PO}$, $\overrightarrow{OX}$, $\overleftarrow{XP}$; and $\overleftarrow{PO}$, $\overleftarrow{OX}$, $\overleftarrow{XP}$. If we confine our discussion to the kind of structures represented in Fig. 6 (i.e., where there is no ambivalence and where all possible positive or negative lines are present), it will be apparent that: when *P* and *O* like each other, the s-digraph is balanced only if both persons either like or dislike *X* (s-digraph *a* is not balanced); when *P* and *O* dislike each other, the s-digraph is balanced only if one person likes *X* and the other person dislikes *X* (s-digraph *b* is balanced); and when one person likes the other but the other dislikes him, the s-digraph must be not balanced (s-digraph *c* is not balanced). These conclusions are consistent with Heider's discussion of *P-O-X* units and with Newcomb's treatment of the *A-B-X* model.

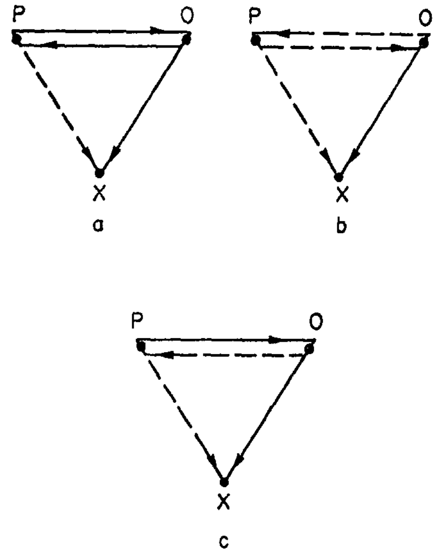

FIG. 6. Three s-digraphs representing Heider's *P-O-X* units. Only structure *b* is balanced.

The further extension of the notion of balance to s-graphs of type 2 remains to be made. The simplest procedure would be simply to ignore the types of lines involved. Then we would again define an *s-graph of type 2* to be *balanced* if all of its cycles are positive. This definition appears to be consistent with Heider's intention, at least as it applies to a situation containing only two entities. For in speaking of such situations having both **L** and **U** relations, he calls them balanced if both relations between the same pair of entities are of the same sign (see **9**, p. 110). There remains some question as to whether this definition will fit empirical findings for cycles of greater length. Until further evidence is available, we advance the above formulation as a tentative definition. Obviously the definition of

[9] In terms of digraph theory we define an *object* as a *point with zero output*. Thus a completely indifferent person is an object. If, psychologically, an impersonal entity is active and likes or dislikes a person or another impersonal entity, then in terms of digraph theory it is not an object.

balance can be given for s-graphs of general type τ in the same way.

SOME THEOREMS ON BALANCE

By definition, an s-graph is balanced if and only if each of its cycles is positive. In a given situation represented by an s-graph, however, it may be impractical to single out each cycle, determine its sign, and then declare that it is balanced only after the positivity of every cycle has been checked. Thus the problem arises of deriving a criterion for determining whether or not a given graph is balanced without having to revert to the definition. This problem is the subject matter of a separate paper (6), in which two necessary and sufficient conditions for an s-graph to be balanced are developed. The first of these is no more useful than the definition in determining by inspection whether an s-graph is balanced, but it does give further insight into the notion of balance. Since the proofs of these theorems may be found in the other paper, we shall not repeat them here.

Theorem. An s-graph is balanced if and only if all paths joining the same pair of points have the same sign.

Thus, we can ascertain that the s-graph of Fig. 7 is balanced either by listing each cycle separately and verifying that it is positive, or, using this theorem, by considering each pair of possible points and verifying that all possible paths joining them have the same sign. For example, all the paths between points A and E are negative, all paths joining A and C are positive, etc.

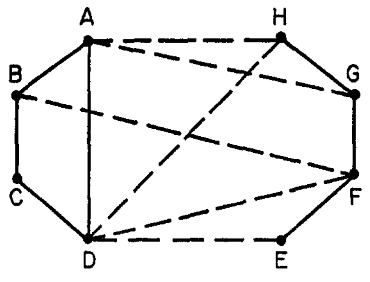

FIG. 7. An s-graph of eight points and thirteen lines which, by aid of the structure theorem, can be readily seen as balanced.

The following structure theorem has the advantage that it is useful in determining whether or not a given s-graph is balanced without an exhaustive check of the sign of every cycle, or of the signs of all paths joining every pair of points.

Structure theorem. An s-graph is balanced if and only if its points can be separated into two mutually exclusive subsets such that each positive line joins two points of the same subset and each negative line joins points from different subsets.

Using the structure theorem, one can see at a glance that the s-graph of Fig. 7 is balanced, for A, B, C, D, and E, F, G, H are clearly two disjoint subsets of the set of all points which satisfy the conditions of the structure theorem.

It is not always quite so easy to determine balance of an s-graph by inspection, for it is not always necessarily true that the points of each of the two subsets are connected to each other. Thus the two s-graphs of Fig. 8 are balanced, even though neither of the two disjoint subsets is a connected subgraph. However, the structure theorem still applies to both of the s-graphs of Fig. 8. In the first graph the appropriate subsets of points are A, D, E, H and B, C, F, G; while in the second one we take A_1, B_1, A_3, B_3, A_5, B_5 and A_2, B_2, A_4, B_4.

In addition to providing two necessary and sufficient conditions for balance, these theorems give us further information about the nature of balance. Thus if we regard the s-graph as representing Heider's **L**-relation in a group, then the structure theorem tells us that the group is necessarily decomposed into two subgroups (cliques) within which the relationships that occur are positive and between which they are negative. The structure theorem, however, does not preclude the possibility that one of the two subsets may be empty—as, for example, when a connected graph contains only positive lines.

The first theorem also leads to some interesting consequences. Suppose it were true, for example, that when two people like each other they can influence each other positively (i.e., produce intended changes in the other), but when two people dislike each other they can only influence each other negatively (i.e., produce changes opposite to those intended). An s-graph depicting the liking relations among a group of people will, then, also depict the potential influence structure of the group. Suppose that Fig. 7 represents such a group. If A attempts to get H to approve of something, H will react by disapproving. If H attempts, in turn, to get G to disapprove of the same thing, he will succeed. Thus A's (indirect) influence upon G is negative. The first theorem tells us that A's influence upon G must be negative, regardless of the path along which the influence passes, since the s-graph is balanced. In general, the sign of the influence exerted by any point upon any other will be the same, *no matter what path is followed*, since the graph is balanced.

By use of the structure theorem it can be shown that in a balanced group any influence from one point to another within the same clique must be positive, even if it passes through individuals outside of the clique, and the influence must be negative if it goes from a person in one clique to a person in the other. (It should be

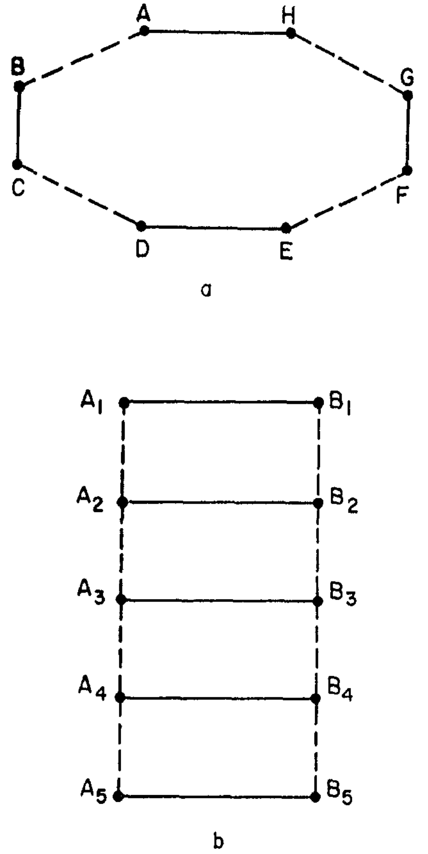

FIG. 8. Two s-graphs whose balance cannot be determined easily by visual inspection.

noted that in this discussion we give the term "clique" a special meaning, as above.) Thus, under the assumed conditions, any exerted influence regarding opinions will tend to produce homogeneity within cliques and opposing opinions between cliques.

Although we have illustrated these theorems by reference to social groups, it should be obvious that they hold for any empirical realizations of s-graphs.

FURTHER CONCEPTS IN THE THEORY OF BALANCE

The concepts of balance as developed up to this point are clearly oversimplifications of the full complexity of situations with which we want to deal. To handle such complex situations more adequately, we need some further concepts.

Thus far we have only considered whether a given s-graph is balanced or not balanced. But it is intuitively clear that some unbalanced s-graphs are "more balanced" than others! This suggests the introduction of some scale of balance, along which the "amount" of balance possessed by an unbalanced s-graph may be measured. Accordingly we define the *degree of balance of an s-graph* as the ratio of the number of positive cycles to the total number of cycles. In symbols, let G be an s-graph,

$c(G)$ = the number of cycles of G,
$c_+(G)$ = the number of positive cycles of G, and
$b(G)$ = the degree of balance of G.

Then

$$b(G) = \frac{c_+(G)}{c(G)}.$$

Since the number $c_+(G)$ can range from zero to $c(G)$ inclusive, it is clear that $b(G)$ lies between 0 and 1. Obviously $b(G) = 1$ if and only if G is balanced. We can give the number $b(G)$ the following probabilistic interpretation: the degree of balance of an s-graph is the probability that a randomly chosen cycle is positive.

Does $b(G) = 50\%$ mean that G is exactly one-half balanced? The answer to this question depends on the possible values which $b(G)$ may assume. This in turn depends on the structure of the s-graph G. Thus, if G is the complete graph of 3 points and G is not balanced, then the only possible value is $b(G) = 0$, since there is only one cycle. Similarly, if the lines of G are as in Fig. 9, some of which may be negative, and if G is not balanced, then the only possible value is $b(G) = \frac{1}{3}$, and $b(G) = 50\%$ does not even occur for this structure. Thus any interpretation given to the numerical value of $b(G)$ must take into account the distribution curve for $b(G)$, which is determined by the structure of G.

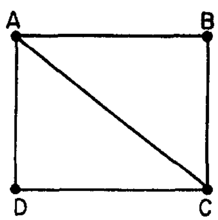

FIG. 9. A graph of four points which can acquire degrees of balance of only .33 and 1.00 regardless of the assignment of positive and negative signs to its five lines.

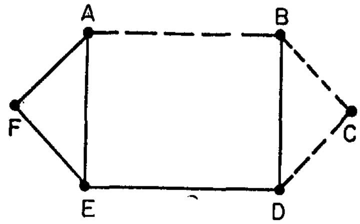

FIG. 10. An s-graph which is 3-balanced but not 4-balanced.

We now consider the corresponding concept of the degree of balance for s-diagraphs. Since an s-digraph is balanced if all of its semicycles are positive, the *degree of balance of an s-digraph* is taken as the ratio of the number of positive semicycles to the total number of semicycles.

In a given s-graph which represents the signed structure of some psychological situation, it may happen that only cycles of length 3 and 4 are important for the purpose of determining balance. Thus in an s-graph representing the relation **L** in a complex group, it will not matter at all to the group as a whole whether a cycle of length 100, say, is positive. To handle this situation rigorously, we define an s-graph to be *N-balanced* if all its cycles of length not exceeding

N are positive. Of course the degree of N-balance is definable and computable for any s-graph. Examples can be given of unbalanced s-graphs which are, however, N-balanced for some N. Figure 10 illustrates this phenomenon for $N = 3$, since all of its 3-cycles are positive, but it has a negative 4-cycle.[10]

For certain problems, one may wish to concentrate only on one distinguished point and determine whether an s-graph is balanced there. This can be accomplished by the notion of local balance. We say an s-graph is *locally balanced at point P* if all cycles through P are positive. Thus the s-graph of Fig. 11 is balanced at points A, B, C, and not balanced at D, E, F. If this figure represents a sociometric structure, then the concept of local balance at A is applicable provided A is completely unconcerned about the relations among D, E, F.

Some combinatorial problems suggested by the notions of local balance and N-balance have been investigated by Harary (7). The principal theorem on local balance, which follows, uses the term "articulation point" which we now define. An *articulation point*[11] of a connected graph is one whose removal[12] results in a disconnected graph. Thus the point D is the only articulation point of Fig. 11. We now state the main theorem on local balance, without proof.

Theorem. If a connected s-graph G is balanced at P, and Q is a point on a cycle passing through P, where Q is not an articulation point, then G is also balanced at Q.

Figure 11 serves to illustrate this theorem, for the s-graph is balanced at A, and is also balanced at B but is not balanced at D, which is an articulation point.

In actual practice, both local balance and N-balance may be employed. This can be handled by introducing the combined concept of local N-balance. Formally we say that an s-graph is *locally N-balanced at P* if all cycles of length not exceeding N and passing through P are positive. Obviously the degree of local N-balance can be defined analogously to the degree of balance.

In summary, the concept of degree of balance removes the limitation of dealing with only balanced or unbalanced structures, and in addition is susceptible to probabilistic and statistical treatment. The definition of local balance enables one to focus at any particular point of the structure. The introduction of N-balance frees us from the necessity of treating all cycles as equally important in determining structural balance. Thus, the extensions of the notion of balance

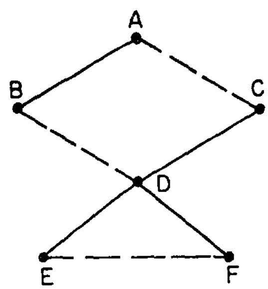

FIG. 11. An s-graph which is locally balanced at points A, B, C but not balanced at points D, E, F.

[10] One way of viewing the definition of N-balance is to regard cycles of length N as having weight 1, and all longer cycles as of weight 0. Of course, it is possible to generalize this idea by assigning weights to each length, e.g., weight $1/2^n$ to length N.

[11] A characterization of the articulation points of a graph, or in other words the liaison persons in a group, is given by Ross and Harary (19), using the "structure matrix" of the graph. An exposition of this concept is given in Harary and Norman (8).

[12] By the removal of a point of a graph is meant the deletion of the point and all lines to which it is incident.

developed in this section permit a study of more complicated situations than does the original definition of Heider.

ADEQUACY OF THE GENERAL THEORY OF BALANCE

In any empirical science the evaluation of a formal model must be concerned with both its formal properties and its applicability to empirical data. An adequate model should account for known findings in a rigorous fashion and lead to new research. Although it is not our purpose in this article to present new data concerning tendencies toward balance in empirical systems, we may attempt to evaluate the adequacy of the proposed general theory of balance in the light of presently available research.

Our review of Heider's theory of balance and of the research findings related to it has revealed certain ambiguities and limitations concerning (*a*) the treatment of unsymmetric relations; (*b*) the generalization to systems containing more than three entities; (*c*) the distinction between the complement and the opposite of a relation; (*d*) the simultaneous existence of relationships of different types; and (*e*) the applicability of the concept of balance to empirical systems other than cognitive ones. We now comment briefly upon the way in which our generalization deals with each of these problems.

Unsymmetric relations. It was noted above that, while theoretical discussions of balance have sometimes allowed for the possibility of unsymmetric relations, no rigorous definition of balance has been developed to encompass situations containing unsymmetric relationships. Furthermore, empirical studies have tended to assume that liking is reciprocated, that each liking relation is symmetric. By stating the definition of balance in terms of s-digraphs, we are able to include in one conceptual scheme both symmetric and unsymmetric relationships. And it is interesting to observe that, according to this definition, whenever the lines $\overrightarrow{PO}$ and $\overrightarrow{OP}$ are of different signs, the s-digraph containing them is not balanced. Thus, to the extent that tendencies toward balance have been effective in the settings empirically studied, the assumption of symmetry has, in fact, been justified.

Situations containing any finite number of entities. Heider's discussion of balance has been confined to structures containing no more than three entities. The definition of balance advanced here contains no such limitation; it is applicable to structures containing any finite number of entities. Whether or not empirical theories of balance will be confirmed by research dealing with larger structures can only be determined by empirical work. It is clear, however, that our generalization is consistent with the more limited definition of Heider.

A relation, its complement, and its opposite. Using s-graphs and s-digraphs to depict relationships between entities allows us to distinguish among three situations: the presence of a relation (positive line), the presence of the opposite of a relation (negative line), and the absence of both (no line). The empirical utilization of this theory requires the ability to distinguish among these three situations. In our earlier discussion of the literature on balance, we noted, however, a tendency to distinguish only the presence or absence of a relationship. It is not always clear, therefore, in attempting to depict previous research in terms of s-graph theory whether a given empirical relationship should be coordinated to no line or to a negative line.

The experiment of Jordan (11) illustrates this problem quite clearly. He employed three entities and specified certain **U** and **L** relations between each pair of entities. The empirical realization of these relations was obtained in the following way: **U** was made into "has some sort of bond or relationship with"; ~**U** into "has no sort of bond or relationship with;" **L** was made into "like;" and ~**L** into "dislike." Viewed in the light of s-graph theory, it would appear that Jordan created s-graphs of type 2 (which may contain positive and negative lines of type **U** and type **L**). It would also appear, however, that the ~**U** relation should be depicted as the absence of any **U**-line but that the ~**L** relation should be depicted as a negative **L**-line. If this interpretation is correct, Jordan's classification of his situations as "balanced" and "imbalanced" will have to be revised. Instead of interpreting the ~**U** relation as a negative line, we shall have to view it as no **U**-line, with the result that all of his situations containing ~**U** relations are vacuously balanced by our definition since there are no cycles.

It is interesting to examine Jordan's data in the light of this reinterpretation. He presented subjects with 64 hypothetical situations, half of which were "balanced" and half "imbalanced" by his definition. He had subjects rate the degree of pleasantness or unpleasantness experienced in each situation (a high score indicating unpleasantness). For "balanced" situations the mean rating was 46 and for "imbalanced" ones, 57.

If, however, we interpret Jordan's ~**U** relation as the absence of a line, his situations must be reclassified. Of his 32 "balanced" situations, 14 have no ~**U** relation and thus remain balanced. The mean unpleasantness score for these is 39. The remaining 18 of his "balanced" situations, having at least one ~**U** relation, become vacuously balanced since no cycle remains. The mean unpleasantness of these vacuously balanced situations is 51. Of Jordan's 32 "imbalanced" situations, 19 contain at least one ~**U** relation, thus also becoming vacuously balanced, and the mean unpleasantness score for these is 51. The remaining 13 situations, by having no ~**U** relations, remain imbalanced, and their mean score is 66. Thus it is clear that the difference in pleasantness between situations classed by Jordan as "balanced" and "imbalanced" is greatly increased if the vacuously balanced situations are removed from both classes (balanced, 39; vacuously balanced, 51; not balanced, 66). These findings lend support to our view that the statement "has no sort of bond or relationship with" should be represented as the absence of a line.[13]

Relations of different types. A basic feature of Heider's theory of balance is the designation of two types of relations (**L** and **U**). Our generalization of the definition of balance permits the inclusion of any number of types of relations. Heider discusses the combination of types of relations only for the situation involving two entities, and it is clear that our definition is consistent with his within this limitation. It is interesting to note

[13] A strict test of our interpretation of Jordan's data is not possible since he specified for any given pair of entities only either the **L** or **U** relation. We can but guess how the subjects filled in the missing relationship. In the light of our discussion of relations of different types, in the next section, it appears that subjects probably assumed a positive unit relation when none was specified, since there is a marked tendency to experience negative liking relations as unpleasant.

that Jordan (**11**) finds positive liking relations to be experienced as more pleasant than negative ones. This finding may be interpreted as indicating a tendency toward "positivity" over and above the tendency toward balance. It is possible, however, that in the hypothetical situations employed by Jordan the subjects assumed positive unit relations between each pair of entities. If this were in fact true, then a positive liking relation would form a positive cycle of length 2 with the positive unit relation, and a negative liking relation would form a negative cycle of length 2 with the positive unit relation. And, according to the theory of balance, the positive cycle should produce more pleasantness than the negative one. This interpretation can be tested only through further research in which the two relations are independently varied.

Empirical applicability of concept of balance. Heider's discussion of balance refers to a cognitive structure, or the life space of a single person. Newcomb suggests that a similar conception may be applicable to interpersonal systems objectively described. Clearly, our definition of balance may be employed whenever the terms "point" and "signed line" can be meaningfully coordinated to empirical data of any sort. Thus, one should be able to characterize a communication network or a power structure as balanced or not. Perhaps it would be feasible to use the same definition in describing neural networks. It must be noted, however, that it is a matter for empirical determination whether or not a tendency to achieve balance will actually be observed in any particular kind of situation, and what the empirical consequences of not balanced configurations are. Before extensive utilization of these notions can be accomplished, certain further conceptual problems regarding balance must be solved.

One of the principal unsolved problems is the development of a systematic treatment of relations of varying strength. We believe that it is possible to deal with the strength of relations by the concept of a graph of strength σ, suggested by Harary and Norman (**8**).

SUMMARY

In this article we have developed a generalization of Heider's theory of balance by use of concepts from the mathematical theory of linear graphs. By defining balance in graph-theoretic terms, we have been able to remove some of the ambiguities found in previous discussions of balance, and to make the concept applicable to a wider range of empirical situations than was previously possible. By introducing the concept *degree of balance*, we have made it possible to treat problems of balance in statistical and probabilistic terms. It should be easier, therefore, to make empirical tests of hypotheses concerning balance.

Although Heider's theory was originally intended to refer only to cognitive structures of an individual person, we propose that the definition of balance may be used generally in describing configurations of many different sorts, such as communication networks, power systems, sociometric structures, systems of orientations, or perhaps neural networks. Only future research can determine whether theories of balance can be established for all of these configurations. The definitions developed here do, in any case, give a rigorous method for describing certain structural aspects of empirical configurations.

REFERENCES

1. BACK, K. The exertion of influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–23.
2. BAVELAS, A. A mathematical model for group structures. *Appl. Anthrop.*, 1948, **7**, No. 3, 16–30.
3. FESTINGER, L. The analysis of sociograms using matrix algebra. *Hum. Relat.*, 1949, **2**, 153–158.
4. FESTINGER, L., SCHACHTER, S., & BACK, K. *Social pressures in informal groups.* New York: Harper, 1950.
5. FESTINGER, L., & THIBAUT, J. Interpersonal communication in small groups. *J. abnorm. soc. Psychol.*, 1951, **46**, 92–99.
6. HARARY, F. On the notion of balance of a signed graph. *Mich. math. J.*, 1953–54, **2**, 143–146.
7. HARARY, F. On local balance and N-balance in signed graphs. *Mich. math. J.*, 1955–56, **3**, 37–41.
8. HARARY, F., & NORMAN, R. Z. *Graph theory as a mathematical model in social science.* Ann Arbor, Mich.: Institute for Social Research, 1953.
9. HEIDER, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, **21**, 107–112.
10. HOROWITZ, M. W., LYONS, J., & PERLMUTTER, H. V. Induction of forces in discussion groups. *Hum. Relat.*, 1951, **4**, 57–76.
11. JORDAN, N. Behavioral forces that are a function of attitudes and of cognitive organization. *Hum. Relat.*, 1953, **6**, 273–287.
12. KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.
13. KÖHLER, W. *Dynamics in psychology.* New York: Liveright, 1940.
14. KÖNIG, D. *Theorie der endlichen und unendlichen Graphen.* New York: Chelsea, 1950 (originally published in Leipzig, 1936).
15. LEWIN, K. *Field theory in social science.* New York: Harper, 1951.
16. NEWCOMB, T. M. *Personality and social change.* New York: Dryden, 1943.
17. NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, **60**, 393–404.
18. PARSONS, T. *Essays in sociological theory.* (Rev. Ed.) Glencoe, Ill.: Free Press, 1954.
19. ROSS, I. C., & HARARY, F. Identification of the liaison persons of an organization using the structure matrix. *Mgmt Sci.*, 1955, **1**, 251–258.
20. SCHACHTER, S. Deviation, rejection and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190–207.
21. TAGIURI, R., BLAKE, R. R., & BRUNER, J. S. Some determinants of the perception of positive and negative feelings in others. *J. abnorm. soc. Psychol.*, 1953, **48**, 585–592.

(Received January 11, 1956)