

THEORY OF SURPLUS IN A MUTUAL
INSURANCE ORGANIZATION

CHARLES L. TROWBRIDGE

THEORY of surplus for a life insurance organization is one among several matters of considerable importance to actuaries that are not comprehensively treated in existing actuarial literature. Although the word "surplus" or its synonyms, with or without modifying adjectives, appear here and there in publications of the Society, with few exceptions the words are used loosely and the implications of surplus are not developed.

Robert Jackson's paper in Volume XI of the *Transactions* is one of the notable exceptions, as is the earlier *Actuarial Studies No. 6* entitled *Distribution of Surplus*, written by Maclean and Marshall. These are classic works intended for students and give an excellent explanation of the apportionment of divisible surplus for ordinary life insurance; they also have something to say about the undivided surplus which remains after the dividend has been declared. Because of the emphasis on the former, however, there is only partial development of the latter.

This paper assumes that the principles of apportioning divisible surplus are well understood but that much remains to be done toward development of a theory of undistributed surplus. Consistent with this assumption, the word "surplus" in this paper will be used in its balance-sheet sense, as the excess of assets over liabilities *after* dividends arising from the current year have been declared.

The paper deliberately leaves to others the theory of surplus for a stock insurance company. In the mutual insurance organization the merging of the customers and the owners into a single group materially simplifies the problem of surplus theory development, since there is one less class of persons whose interests must be considered. This paper concerns itself only with the surplus of the mutual life insurance company, but the development is not restricted to any particular line of the company's activity. Indeed, there is little here that could not be applied to other insurance organizations operating under mutual principles, and much of the development would seem to apply to mutual financial institutions not a part of the insurance industry.

This paper deliberately emphasizes the development of theory and leaves aside many of the practicalities. No implication is intended that

practical considerations are unimportant; it is felt, however, that good practical solutions will be difficult until a theoretical structure has been built. Even if a clear-cut and tidy theory of surplus can someday be developed, practice would undoubtedly diverge for any one of several reasons, not the least of which is the competitive pressure exerted by the free-enterprise system. No doubt the difficulty of fitting practice to theory accounts to a large extent for the paucity of the latter, but it should not deter the theorist from making the attempt.

It is the premise of this paper that a satisfactory surplus theory is vital to the good management of a mutual insurance enterprise. Surely the "contribution to surplus" is one of the important elements in the pricing of any insurance product, and it is directly related to financial solvency as well. Sound pricing theory and the financial solvency of the company are among the prime responsibilities of the insurance company actuary, who must therefore operate under some surplus theory whether he thinks of it this way or not.

It would seem that a comprehensive theory of surplus should face up to at least these important questions: What exactly is surplus, what is its purpose, of what magnitude might it be, from whence does it arise, and what are its implications with respect to equity between policyholders? These, particularly the last, are weightier questions than they first appear to be.

This paper is an attempt to pose the important questions, to make at least a start toward their solution, and therefore to develop the beginnings of a theory of surplus.

WHAT IS SURPLUS?

Some students of this subject may be satisfied to answer this first question from the framework of the published annual statement. If one is willing to accept for surplus-theory purposes the methods and assumptions underlying the valuation of statement assets and liabilities, the sum of all amounts appearing below line 26 of page 3 of the NAIC statement of the mutual life insurance company can be viewed as the measure of surplus. Note that this definition would include those of Mr. Jackson's "contingency reserves" appearing below line 26.

We should recognize that surplus is essentially undivided and unallocated, despite the tendency to hold surplus under several different labels in the NAIC statement. If, for example, solvency were to be threatened by some mortality catastrophe, a contingency reserve bearing an investment label would probably be as effective as one bearing a mortality name in keeping the company afloat. To think of contingency reserves as

multiple rather than singular seems unnecessarily confusing. This point is an echo of Mr. Maclean's discussion of the Jackson paper. Many would agree with Mr. Maclean that any items in the nature of a contingency reserve, particularly the Mandatory Security Valuation Reserve, which appear above line 26 should also be treated as a part of surplus.

Others may not be satisfied to define surplus exactly as the statement does. They may point to the conservatism commonly and legitimately employed for statement purposes, in the net premium valuation methods generally in use in North America, the conservative actuarial assumptions typically used, and the generally conservative valuation of assets. These persons may prefer to value assets and liabilities more realistically (not for statement purposes necessarily but for the purposes of surplus theory) in order to concentrate *all* elements of conservatism into the surplus item toward which the theory is directed.

Some might go so far as to define surplus in terms of assets and liabilities valued on best-estimate assumptions and prospective gross-premium-valuation methods, recognizing rates of lapse, surrender, election of options, and expense, and possibly provision for future dividends, as well as the more usual mortality, morbidity, and interest.

Actuaries who may be appalled at the amount of effort implied for this second calculation of assets and liabilities, or who prefer other bases in connection with surplus definition, will find that the discussion of other surplus questions following does not hang on the particular definition of surplus chosen.

WHAT IS THE PURPOSE OF SURPLUS?

There is a clear primary reason why mutual companies, almost without exception, hold rather substantial surpluses, and add to them in most years, in preference to paying 100 per cent of "earnings" out as "dividends." In its primary role surplus is a reserve against the several contingencies which could impair the ability of the company to meet its obligations.

The contingencies against which the surplus is a hedge can be viewed as many and varied—or they may be viewed as special cases of the two best-recognized hazards: (1) the possibility of loss from asset depreciation and (2) the possibility of loss from inadequate premium. Within the former are the possibilities of default on debt obligations, of decline in the market values of stocks, of economic damage to real estate, and of major physical destruction of assets or the security behind assets as a result of war. Within the latter are the hazards of epidemic, of air or water pollution, of war or other catastrophes affecting the health of

many, of inadequate information on which to base the premium, as well as the possibilities of inflation or other economic or social factors affecting the cost of the insurance product or the price at which it can be sold.

Some loss of each of the two types is probably unavoidable if a company operates at all widely over any substantial time period; but many of these losses turn out to be temporary and recoverable. The depressed asset value recovers or the poor mortality experience is reversed or is later offset by increased premium or reduced dividends. Any substantial threat to solvency is relatively unlikely in a well-established, conservatively managed company in this country—but it is nonetheless the *possibility* of severe adverse development, coupled with the serious consequences for policyholders if insolvency should occur, that justifies the holding of a modest surplus.

There is another role for surplus that is not as well recognized as the contingency reserve role. Although a contribution to surplus is commonly thought to increase the price paid by the customer-owner for his insurance, and although this belief is probably correct under most theories of surplus during the early period of a company's growth, eventually the reverse is true and the presence of a surplus actually decreases the cost of insurance. Surplus generates investment income over and above the earnings on the assets covering the liabilities, and this additional investment income is available to increase the dividends or otherwise lower the cost of the product. It is easily shown that the net contribution to surplus (from the entire body of policy-owners) is negative in any year that surplus grows at a lesser rate than the rate of investment earnings. Surplus thus has a role in a company's competitive position, by improving it once the surplus is largely built even as it hurts it during the period that surplus is being accumulated.

Another form of the competitive role of surplus lies in its ability to increase the company's power to take advantage of opportunities that may present themselves. A company desiring to expand into a new line of insurance products, or into a new marketing territory, can more readily do so if it has a substantial surplus from which to finance the new endeavor in its development period. A company can take a bolder approach to investment opportunities or to underwriting or pricing decisions if assets are well in excess of liabilities. In short, surplus increases management's range of choice and should therefore be a positive factor toward its position in the market place. In this sense the surplus plays a role similar to the "net worth" of the typical industrial corporation.

Despite the importance of these competitive considerations in the de-

velopment of a surplus theory, it is likely that the primary function of surplus remains that of a contingency reserve or, at least, that a surplus developed along contingency reserve lines can be designed to serve competitive considerations as well. We will view surplus largely in that context as we proceed to the important question of how large surplus might reasonably be.

SURPLUS OF WHAT MAGNITUDE?

The traditional measure of the adequacy of surplus in life companies has been the ratio of surplus to liabilities, or sometimes surplus to assets. Obviously the investment risk has been emphasized. For other insurance lines operating largely on a term basis, where both assets and liabilities are small in relation to the premium risk assumed, the traditional measure of surplus is its ratio to premium, the emphasis here being on the insurance risk. Since the insurance products sold by life companies today involve substantial proportions of group life, group accident and health, and ordinary insurance or reinsurance written on a term basis, a satisfactory measure of surplus adequacy would seem to involve premiums as well as assets, the insurance risk as well as the investment risk.

Recognizing that the hazards against which the surplus is a hedge are of more than one type, the mathematical form of a reasonable surplus goal may thus have more than one component. As perhaps the simplest mathematical form to express a goal for a contingency reserve against n hazards, we might use

$$S = \sum_1^n f_k P_k,$$

where S is the surplus objective; P_k is a parameter chosen as a measure of hazard k ; f_k is the fraction of P_k that defines the k th component of S . For example, we might choose to recognize only two risks—the asset-depreciation risk proportional to $P_1 =$ assets and the pricing risk proportional to $P_2 =$ premium. Then $S = f_1 P_1 + f_2 P_2$. In a typical life company P_2 would likely include only the pure insurance or risk portion of the premium (with any savings element ignored), and this will *always* be assumed in the development that follows. f_1 and f_2 would be set, both as to absolute and as to relative size, to suit the actuary's view as to the magnitude of S and as to the relative weight to be assigned to each of the two contingencies.

Starting from this simple two-term linear compound concept, as many refinements can be added as the theorist feels important. Among the refinements that may appeal to some are these:

1. Additional components might be added to reflect other hazards not included within the two already indicated.

2. Assets might be subdivided by type of investment, each with a separate f_1 , on the rather solid grounds that the risk of asset depreciation varies with the mix of types of investment.

3. Similarly, P_2 , the risk premium, might be subdivided by type of coverage, each with a separate f_2 , on the grounds that relatively high participating premiums are less likely to prove inadequate than premiums close to a nonpar level or that premiums guaranteed for a short time involve less risk than those contractually guaranteed for the life of the policy. The same differentiations could generally be accomplished by use of a single f_2 but with a more sophisticated definition of risk premium. It may be felt important somehow to work the mortality risk on annuity coverages into the risk-premium concept.

4. The f 's might be viewed as variables, decreasing as the P 's grow, on the grounds that a company of larger size and wider spread of risk needs a smaller surplus (relatively) than it did when it was smaller.

Despite the validity of some of these complicating factors, the author warns against overrefinement on the practical grounds that there is little of science and much that is arbitrary in setting the f 's. He has the feeling that somewhat oversimplified concepts may here prove to be more satisfactory.

Although actuaries should be able to reach reasonable agreement on a mathematical *form* for a surplus objective, either in the simple two-parameter form suggested earlier or in some more refined form of the linear compound concept, it is highly unlikely that any two actuaries approaching the problem independently would arrive at the same f 's. This is because the actuary has little solid statistical theory on which to base an answer to the question of surplus magnitude, since the surplus is intended to do more than cover random statistical fluctuation. This paper will therefore make no attempt to answer the "how much surplus" question from a theoretical point of view but will instead attempt to determine what surpluses various types of institutions do in fact hold as hedge against the two most apparent contingencies.

A problem one immediately faces is that life companies have two major kinds of risk, and the surplus that a life company holds is a contingency reserve against both. The analysis would be much simpler if each could be viewed separately. To a degree this may be possible if we can discover other institutions organized on the mutual principle which face these hazards separately.

First, let us concentrate on risks of an investment nature—primarily that of asset depreciation but perhaps also related to interest guarantees. For comparison we immediately think of mutually organized savings and

loan institutions. These institutions invest funds in the long-term mortgage market, and their risks are almost entirely of an investment nature. These are obviously not identical to the similar risks in a life company, but they have many of the same characteristics. Much the same, perhaps, could be said with respect to mutual savings banks. Certainly the extent to which these institutions hold surpluses, which must be largely in the nature of a "contingency reserve for investment loss," may have some bearing on the question before us.

Table 1 presents recent data as to statement surpluses as a percentage of assets in mutually organized savings and loan institutions and mutual savings banks. Whether there may be "hidden" surplus in these ratios, the author has too little knowledge to evaluate—nor does he attempt to assess the relative investment risk between each of these two types of financial institutions and the life company portfolio.

We would also do well to examine mutual organizations whose risk is essentially of an insurance nature—insurance companies or associations operating largely on the term principle. Here we think of the mutual property and casualty companies, mutual accident and health companies, and the Blues. Table 2 shows recent data on the relationship between surplus and yearly premium in these various organizations. Once again,

TABLE 1
STATEMENT SURPLUS AS A PERCENTAGE OF ASSETS

	Per Cent
Savings and loan organizations	6.7*
Mutual savings banks	7.7†

* From *Savings and Loan Fact Book, 1966*: Aggregate ratio of all (over 6,000) United States savings and loan organizations.

† From *Moody's, 1966*: Unweighted average of ratios of a sample of 40 of the largest 100 (10 largest plus every third of next 90).

TABLE 2
STATEMENT SURPLUS AS A PERCENTAGE
OF YEARLY PREMIUM

	Per Cent
Mutual fire and casualty companies	45*
Mutual accident and health companies	36†
Blue Cross associations	19‡
Blue Shield associations	27‡

* Statement figures as of 12/31/65 from *Best's Insurance Reports*. Median of ratios of 32 large mutuals.

† Statement figures as of 12/31/65 from *Best's Insurance Reports*. Median of ratios of 11 mutuals, all that could be identified as primarily accident and health companies.

‡ From 1966 *Argus Health Chart*. Aggregate ratio of all Blue Cross associations and separate aggregate ratio of all Blue Shield associations. A few associations that are both are included in both aggregates.

no judgment is offered as to degree of risk in the various premium mixes or as to the realism in the surplus figures shown. In calculating the ratios shown in Table 2, an attempt was made to include as surplus all voluntary or contingency reserves in the nature of surplus but to exclude all reserves in the nature of liabilities.

Finally, there is some point in trying to evaluate the extent to which the surplus goals for the two rather different risks are additive. Put more specifically, if it were felt that an investment institution needed a surplus of h_1P_1 and that a pure insurance company needed h_2P_2 , what would be a reasonable goal for a life insurance company subject to *both* investment and insurance hazards? The extreme conservative viewpoint might well be the sum of the two separate surplus goals, emphasizing the chance that unfortunate developments can happen simultaneously in both the investment and the insurance operations. The most optimistic viewpoint is toward a surplus goal that is simply the greater of h_1P_1 and h_2P_2 , emphasizing the idea that surplus against one hazard can be used to meet another. A midground viewpoint is that an appropriate goal is $f_1P_1 + f_2P_2$, such that the resulting S is greater than the larger of h_1P_1 and h_2P_2 but smaller than their sum. Greater precision in the theory of the combination of two risks may be rather meaningless in view of the rudimentary state of our knowledge as to the measure of the risks viewed separately.

SOURCES OF SURPLUS

A first source of surplus is interest on surplus already built. A second is net capital gains, if the Mandatory Security Valuation Reserve is considered to be within the definition of surplus or if its required level has already been reached. Despite these two possibilities, the main source must be that part of the earnings from insurance and normal investment operations not paid out in dividends. This contribution to surplus can be the result of a direct charge for surplus in the dividend calculation or it can come about indirectly by using somewhat conservative claim, interest, and expense factors in the dividend formula.

Assume that a company is "on target" at the beginning of a year, such that $S = f_1P_1 + f_2P_2$. To remain on target at the end of the year, the contribution to surplus ΔS for the year must then be

$$\Delta S = f_1\Delta P_1 + f_2\Delta P_2 = g_1f_1P_1 + g_2f_2P_2,$$

where g_1 is the growth rate as to assets and equals $\Delta P_1/P_1$ and g_2 is the growth rate as to risk premium and equals $\Delta P_2/P_2$.

For many purposes the first difference of the surplus goal ΔS is more meaningful than the surplus goal itself, because it is subject to better

control and better relates to the current situation. For this reason it will be found that much of the following development emphasizes ΔS rather than S .

The obvious first source for ΔS is the investment earnings on surplus itself, $iS = if_1P_1 + if_2P_2$. There still remains $\Delta S - iS = (g_1 - i)f_1P_1 + (g_2 - i)f_2P_2$ to be found elsewhere. In the calculation of dividends, a direct or indirect charge or holdback of $(g_1 - i)f_1P_1 + (g_2 - i)f_2P_2$ would therefore seem to be necessary to keep S at the intended level; and the investment earnings on previous surplus must be reserved for this purpose as well.

In a life company in a Phase 1 income tax position, it is not uncommon to think of the investment rate i in two ways—a “before income tax” rate and a lower “aftertax” rate. In the foregoing analysis either is satisfactory, providing federal tax is consistently handled in the dividend calculation. If i is the “before tax” investment rate i' , all federal income tax must be charged in the dividend calculation. If i is an “aftertax” rate i'' for interest on surplus, only federal tax in excess of $(i' - i'')S$ need be provided for in the dividend calculation.

This analysis of ΔS clearly shows how sensitive a year's surplus requirements (other than interest on surplus itself) are to the growth rates g_1 and g_2 . As long as the parameters in terms of which the surplus goal is expressed are growing at a rate higher than the investment earning rate, $g - i$ and $\Delta S - iS$ are positive, and the contribution to surplus represents a reduction in dividends and an increase in price. This is no doubt the present situation in which most companies find themselves, but not all to the same degree. If i is 5 per cent and all other things are equal, a company with a 10 per cent growth rate on all parameters considered would require \$5 of surplus contribution (other than interest on surplus itself) for every \$1 needed by its competitor with a 6 per cent growth rate.

Although companies expect to grow at a rate faster than the interest rate, and most of them currently do, we complete the analysis only if we recognize that g can become less than i —in fact, g can become zero or even negative. Indeed, unless a company is presumed to have a perpetual and ever growing existence, $g - i$ can be expected to be negative in the company's mature or declining stages. Whenever $g < i$, the contribution to surplus turns into a contribution from surplus and serves to reduce the price of insurance rather than to increase it. This is a common practical situation with respect to certain subdivisions of life companies, even though it may not yet be with respect to the entirety of any company.

There can be the in-between position, where one parameter (e.g., premium) is growing faster than the interest rate, whereas another (e.g.,

assets) grows slower than interest. In this circumstance, $\Delta S - iS$ would have both a negative and a positive component, the sign of the sum depending on the relative sizes of the g 's, the f 's, and the parameters themselves. Life companies have tended toward slow asset growth in recent years; a few of them have paid out more in benefits, expenses, and dividends than they have received in premiums; in these cases assets obviously grow at a lesser rate than the investment earnings rate.

SURPLUS THEORY AND ITS RELEVANCE TO EQUITY

Among the major obligations of the actuary of a mutual insurance organization are the establishment and preservation of equity in the pricing of various insurance products to several generations of policyholders. Clearly, equity cannot be specifically defined to the satisfaction of all, much less exactly achieved in a competitive world. It is also clear that surplus considerations are only a part, perhaps a relatively minor part, of the over-all equity problem. Equitable allocation of interest earnings, of expenses, or of mortality-morbidity experience may have more impact on pricing than equitable allocation of surplus contribution.

Despite everything said in the preceding paragraph, any theory of surplus must have answers to the not-unimportant questions relating to which policyholders are asked to contribute how much toward surplus and to the related question of whether surplus is returned when a policyholder leaves the group and is no longer a part of the risk against which surplus is held.

We can point up the problem involved by contrasting the typical corporation organized for profit-making purposes with the mutual life company. In the former the excess of assets over liabilities becomes the equity or net worth, which belongs to the owners of the enterprise—the stockholders. Equity between succeeding generations of stockholders is preserved, even though all earnings are not immediately paid out as dividends to stockholders, by the changing price for the common shares bought and sold in a free market. In the mutual insurance organization there is no similar mechanism automatically preserving equity between generations of policy-owners, since the terminating policyholder cannot sell his share in the company surplus. This lack of a marketable ownership share is one factor pointing up the importance of the issues raised in the preceding paragraph.

We shall see that the theory of surplus so far put forth here can logically lead to more than one set of answers to these equity questions, their essential differences lying in the degree of pooling or the extent to which policyholders are viewed together for surplus contribution purposes. This

should not be surprising, since the degree of pooling is often the key point in other considerations of equity—the assessment of expense, the distribution of investment earnings, or the allocation of claims. It might well be pointed out here that the following discussion assumes that the actuary is satisfied that principles of equity have been followed with respect to allocation of expense, interest, and claims and that only the surplus contributions are yet to be examined in establishing the pricing structure.

The Policy-based Contribution to Surplus

To this point the surplus goal has been stated in terms of an insurance company as a whole. If we adopt the concept that each policy (ordinary, industrial, or group) should provide surplus in proportion to its portion of the over-all need for surplus, we can apply the same criteria policy by policy. Then each policy r is expected at any time to have built up a surplus defined by $S^r = f_1 P_1^r + f_2 P_2^r$, where P_1^r and P_2^r are the assets and the risk premium of that particular policy, respectively. The whole is equal to the sum of its parts, and the sum of the individual surplus goals equals the over-all surplus goal.

If we now can assume that a policy is on target as to surplus at the beginning of a year, then, for that policy to remain on target at the end of the year, it must contribute to surplus $\Delta S^r = g_1^r f_1 P_1^r + g_2^r f_2 P_2^r$. The form of this function is exactly the same as the similar function appropriate to the company as a whole, but g_1^r and g_2^r are now the growth rates for the parameters of the particular policy, not the corresponding growth rates for the whole company. The g^r 's are much more variable than the rather stable growth rates of the company as a whole, and have some interesting characteristics at the two extremes, when a policy is first issued and when it ultimately goes off the books.

At the beginning of the year during which a policy is issued, its parameters are zero, the surplus it has built up is zero, and the policy can be said to be on target. At the end of that year $S^r = f_1 P_1^r + f_2 P_2^r$ should be built up, where the P 's are the values of the parameters for the particular policy at the *end* of the year. Therefore, the contribution to surplus from policy r for the initial year is $f_1 P_1^r + f_2 P_2^r$. This formula does not look very much like the formula in the previous paragraph, which has technical difficulties in this initial year because it is expressed in terms of the zero parameters at the *beginning* of the year, and an infinite growth rate g . The above form corrects these difficulties. It is clear, in any case, that the initial surplus contribution is substantial, no doubt higher than can be practically attained.

For the year in which the policy goes off the books, the growth rates become -1 , and the net effect is to return to the policy (probably in the form of a terminal dividend) all previously contributed surplus. The terminal-dividend or returnable-surplus concept is entirely consistent with the underlying concept upon which the policy-based contribution to surplus is conceived. The long-range surplus charge for any policy is nil; in the early years, however, the substantial holdback for surplus has a noticeable effect on the cost of insurance for those early years.

This policy-based approach has as its strongest point a rationale that seems to be the ultimate in individual equity. It also results in a long-range pricing policy which can truly be described as "insurance at cost." On the other hand, it has difficulties, some of which are enumerated below:

1. The whole concept of individual returnable surplus is in many ways foreign to the co-operative principle upon which mutual insurance is built. No policy gets help from or gives it to any other policy, so far as surplus considerations are concerned.

2. In the short term, dividends otherwise payable are hit hard by the necessity of rapid surplus accumulation. This is likely to make a company noncompetitive from a price point of view, particularly in the typical type of short-range net cost comparison.

3. Unless a similar theory has been followed since the company was established, it would be highly unlikely that the present surplus position of each policy could be calculated. Even if this were possible, it would be found that many policies are "off target." Theoretically, adjustments could be made, but the immediate effect on dividends would be troublesome, and the whole concept could easily die because of the practical difficulties by which it is surrounded.

The Segment-based Contribution to Surplus

If we ignore the single policy as a unit of surplus analysis and direct our attention instead to a group of policies with similar characteristics, we modify the theory away from the policy-based concept to what we might call the segment-based theory of surplus allocation. Define a segment as any group of policies which are to be considered together for surplus purposes. Perhaps one segment might be all policies written over a particular time period at a particular age on a particular policy form; or it might be the entirety of a particular line—for example, individual health insurance. There are altogether n segments making up the whole, where n theoretically lies anywhere greater than 1 and less than the total number of policies. It is only necessary that each policy—past, present, or future—have a home within one and only one segment. Some segments might be closed groups (no new policies joining), others open groups. The

concept clearly permits new segments to come into existence in the future and some old segments to go entirely out of existence.

We now define the surplus goal for any segment s as $S^s = f_1 P_1^s + f_2 P_2^s$. ΔS^s then becomes $g_1^s f_1 P_1^s + g_2^s f_2 P_2^s$, and $\Delta S^s - iS^s = (g_1^s - i)f_1 P_1^s + (g_2^s - i)f_2 P_2^s$.

This form is by now familiar, but P_1^s and P_2^s now represent the parameters *for the segment*, and g_1^s and g_2^s represent the growth rates of P_1^s and P_2^s . These growth rates would not normally be the same as the g 's for the company as a whole or as the g^r 's representing the growth rate of any single policy within the segment. The g^s 's are, of course, a weighted average of the g^r 's within the segment, and the g 's are a weighted average of g^s 's.

The growth rates g^s , while more stable than the g^r 's and less likely to go negative, are still less stable than the g 's for the company as a whole. Moreover, growth rates can be expected to vary widely from segment to segment. The annual contribution to surplus expected from a fast-growing open segment, whether expressed in the ΔS or the $\Delta S - iS$ form, would be higher than from a slow-growing mature segment, while any decreasing segment (particularly a closed segment) would be paying out surplus accumulated in the past. As long as a segment remains open for new business, any surplus built up from that segment in the past is available to help meet the surplus objectives of all policies entering that segment. A new segment gets no help at all, however, and a closed segment only takes care of its own.

For any segment not composed of future business only, it could well be difficult to establish its level of surplus as of the date the theory became effective, and one would expect to find the level "off target" in many segments. If so, some adjustments might be called for in order to move toward the target. One straightforward way is to expect annual contributions toward surplus in accordance with $\Delta S^s = g_1^s f_1 P_1^s + g_2^s f_2 P_2^s$ but to modify the g 's or the f 's upward if surplus is behind and downward if ahead.

Note that the theory here calls for an annual addition to surplus from the entire segment of $\Delta S^s = g_1^s f_1 P_1^s + g_2^s f_2 P_2^s$ (perhaps with modified g 's or f 's); but it is silent with respect to the way in which each policy r within the segment contributes. As long as the theory does not speak to the question of the allocation of the surplus objective within the segment, it is of little use in the determination of equitable premiums and dividends. We can fill this vacuum in a manner consistent with the underlying assumption that the segment is the unit for surplus allocation by emphasizing ΔS^s and allocating it to policies within the segment strictly in

accordance with $g_1^i f_1 P_1^r + g_2^i f_2 P_2^r$. The effect is that for any particular policy one abandons the accumulated surplus test based on S and substitutes therefor the annual contribution to surplus test based on ΔS .

At first glance this appears to be a throwback to the policy-based theory, but it is not the same. The g 's are those for the segment (not the single policy), so there is socialization as to rate of growth. This is *not* a terminal-dividend theory releasing surplus to the individual policy when its own parameters go to zero; instead it is a release of surplus to the *survivors* of the segment as the parameters for that segment begin to decrease. It has a tontine effect, with especially good performance for the last few survivors of the segment.

Alternatively, the contribution to surplus test for a particular policy within the segment could be based on $\Delta S^s - iS^s$ rather than on ΔS . If the iS^s portion of ΔS^s is first credited to the segment, and then in addition each policy is expected to contribute $(g_1^i - i)f_1 P_1^r + (g_2^i - i)f_2 P_2^r$, we find a pooling with respect to the current surplus position in addition to the pooling with respect to the growth rates. Fast growing (new) policies get a slightly more favorable treatment under this approach, because they get help from interest on a segment-average surplus rather than on the surplus that each policy has been able to build up by itself. In other respects this concept of an equitable contribution to surplus from any policy within the segment is largely similar to the form of segment-based contribution to surplus based on ΔS .

The Company-based Contribution to Surplus

We have seen that the segment-based surplus allocation principles involve a first step in the socialization or pooling of the policy-based principles. If the number of segments is large and there are few policies within each segment, segment-based principles are only a small step removed from our policy-based starting place—but as the number of segments decreases and the number of policies within each segment grows, the gap widens. Carrying this process farther, we might define one segment as the entire ordinary business, a second as the entire group business, and the third as the whole of the industrial business of a company. The ultimate step is to think in terms of one open segment—the entire business of the company, past, present, and future. This leads to the company-based contribution to surplus.

The principles are easily derived from the segment-based principles already established. There is only one s , and P_1^s , P_2^s , g_1^s , and g_2^s become P_1 , P_2 , g_1 , and g_2 , respectively. The pooling with respect to growth rates is complete. Equity (from a surplus point of view) for any policy r or

segment s is defined by a surplus contribution (including interest on surplus already built up) equal to $g_1 f_1 P_1^{r \text{ or } s} + g_2 f_2 P_2^{r \text{ or } s}$, where the g 's are the same for all policies or segments and are based on the over-all company growth rates. If we prefer, we can pool with respect to current surplus position as well, in which case we ask each policy or segment to contribute to surplus (in addition to interest on surplus already built up) $(g_1 - i) f_1 P_1^{r \text{ or } s} + (g_2 - i) f_2 P_2^{r \text{ or } s}$.

Surplus thus is returnable only if the g 's go negative. Some part of the interest on surplus is returnable if the g 's fall below i . As long as g exceeds i , a small surplus contribution is expected from each and every policy and surplus is viewed as nonreturnable.

Note that under this single-segment concept a declining industrial business might be releasing surplus to an expanding ordinary business, whereas the contrasting two-segment approach would pay out past surplus built up from industrial policyholders to the survivors of the industrial business, meanwhile expecting the expanding ordinary to build up a surplus entirely from its own resources.

Summary of Three Approaches

Although three approaches to surplus allocation have been discussed as if each were different and distinct, the most general of these approaches really includes the other two. The segment-based approach with n segments grades into the company-based concept of returnable surplus as n increases toward the total number of policies. In the other direction, it grades into the policy-based concept of nonreturnable surplus as n decreases toward unity. Between these two extremes the division of the total set of policies into subsets or segments can be done in an almost infinite number of ways. One of these ways may be more logical than another, but there is little reason to say that one is more equitable than another. The policy-based approach is tenable from an equity viewpoint; but so is the company-based approach or most segment-based approaches between.

The distinction between even the extremes of the continuum are subtle. It is to be noted, for example, that the company-based contribution described above as nonreturnable is only relatively so. If the company ever were to go into a long-term decline with respect to the parameters on which the surplus objective is based, surely surplus should be released through the pricing structure. This could serve to restimulate the company's growth, but, if not, the surplus must obviously go back to the remaining policyholders, since there is no other place for it. Even if the company continues indefinitely to grow faster than i , surplus can prove

to be returnable in another sense if any of the contingencies against which the surplus is held should actually take place.

If Surplus Declines

To this point the discussion of the building toward surplus goals has assumed that the process is not interrupted by the occurrence of the unfavorable events against which the surplus is held. To round out the discussion of equity between policyholders in the surplus area, we must examine the possibility that surplus will decline or even vanish through the occurrence of some unexpected loss. Such loss may result from investments, from insurance operations, or from a combination of both.

In the year of the unexpected unfavorable event or events, surplus will actually decline if the resulting loss is greater than the contribution to surplus that would otherwise have occurred. Even if the absolute size of surplus does not diminish, the loss may cause a failure to meet surplus objectives. With respect to the company as a whole, there is little alternative but to let the surplus perform its function of absorbing the loss and then begin a surplus rebuilding program.

The question remains as to how much of the additional contribution needed to rebuild surplus is to come from any particular segment or any particular policy. The actuary's view as to this question is likely to depend on the circumstances surrounding the loss, particularly if the loss appears to be especially attributable to certain policies or groups thereof.

As one example, assume a substantial capital loss due to massive default in debt obligations. Generally speaking and with some notable exceptions, life companies view invested assets as pooled, with each policy or segment "owning" its proportionate share of all invested assets rather than an interest in particular investments. This pooled-investment concept would likely lead to a spreading of the capital loss to policies or segments in proportion to the P_1 or σ . This is equivalent to rebuilding surplus in the same manner.

By way of contrast, assume a substantial loss due to the underpricing of income disability riders attached to ordinary policies. Similar to the first example, the surplus rebuilding necessary after the loss could be assessed against each policy or segment in proportion to P_2 or σ . This would be the likely result, however, only if the insurance risk were treated as pooled across policy and segment lines. When this actually did occur back in the 1930's, the losses on income disability riders were assessed in most companies against only those segments exposed to the particular disability risk. Companies that did this were upheld by court decisions.

It seems clear that equity in surplus once a substantial loss has oc-

curred becomes another form of equity in the allocation of investment results and claim experience.

CONCLUSION

The questions with respect to surplus raised in the introductory portion of this paper have not been entirely answered. Controversy has surrounded and will continue to surround the questions of how much surplus, of returnable vs. nonreturnable surplus, and perhaps even as to how surplus is defined or what its purpose is.

There may well be other important questions which have not been faced here. Some of these lie in the practical rather than the theoretical area, an area that has been deliberately de-emphasized. Thus, for example, we have made no reference to such practical considerations as statutory restrictions on surplus or court decisions affecting surplus accumulation and distribution. Still other questions not considered by the paper may go to the heart of the objectives of a mutual company. If, for example, growth is a legitimate objective of a mutual insurance organization (and there seems to be little reason to question the prevailing opinion that the organization must grow to preserve its vitality and to serve its members better), then surplus theories of the company should encourage growth, or at least not discourage it. The reader may wish to examine the theory developed here from that point of view.

It is to be hoped that actuaries of mutual companies will be stimulated to add to or subtract from the concepts here expressed and that actuaries representing stock companies will see fit to develop the necessary modifications to reflect the different organization of the stock company.

DISCUSSION OF PRECEDING PAPER

ROBERT T. JACKSON:

Mr. Trowbridge is to be congratulated on attacking a very interesting, if tantalizing, subject to the actuary. Nowhere else, so far as I know, is there anything like the comprehensive treatment of surplus itself—its source, magnitude, and uses—as that given in this paper. Most others who have written on the subject at all have done so solely as an appendage to a discussion of surplus distribution. That we have had to wait so long for a scholarly analysis of surplus itself may, I suspect, be due to the fact that the subject is as elusive in concrete terms as it is fascinating in theory.

One of Mr. Trowbridge's interesting points is a comparison of the surplus held by insurance companies with those held by other financial institutions. Although, as he says, it is very difficult to compare needs for surplus, it would seem to me that the insurance companies as a whole are probably carrying a somewhat larger surplus than mutual savings banks and savings and loan associations. There is usually a degree of conservatism, particularly during a higher interest-rate period, in our liabilities which creates a margin over and above pure surplus for most companies. Although, like Mr. Trowbridge, I know too little about the operation of savings banks to be able to offer a positive conclusion, it is difficult to see that there could be comparable margin for conservatism in a mutual savings bank. Of course, we are exposed to the hazard of premium deficiency while the savings bank is not—a sufficient reason to argue that our surplus position is not necessarily more conservative than that of these institutions.

This brings us to the meat of the problem—the proper amount of surplus, which obviously is tied into its potential uses. It seems clear to me that we must consider the amount of surplus to be carried within a frame of reference limited to those likely catastrophes under which we can expect to continue to exist as a business. This, in my mind, rules out, for example, a major atomic war, for which no amount of surplus would suffice.

For somewhat the same reason, I question whether excessive operating costs—always mentioned but never dwelled on in the orthodox treatise on the subject—are, in fact, a hazard against which surplus must be held. Surely our other margins must be large enough to absorb creeping inflation through dividend reduction, while galloping inflation would doubtless render our guarantees so unattractive that we would face wholesale

surrender at a time when our fixed-income-bearing assets were so seriously impaired in market value that no reasonable surplus position could avert catastrophe.

I am suggesting that surplus must be viewed as a hedge against market fluctuations, within reason, and mortality fluctuations, within reason, and that determination of an appropriate surplus objective will be hopelessly bogged down if it is not clearly recognized that no surplus can be sufficient to preserve the company against every possible catastrophe.

Further, there are two imponderables in the event of a serious catastrophe which also affect the "proper" amount of surplus—the public's degree of confidence in the life insurance industry and the individual company (which could be of critical importance in the earlier stages of a catastrophe, as in the Great Depression) and, at some point, if the entire industry is threatened, the virtual certainty of government intervention in an attempt to preserve the industry.

Mr. Trowbridge poses a very valid question, without giving us answers: If one surplus amount is required for one contingency—mortality—and another for a second contingency—asset depreciation—is the total surplus properly the sum of the two, the larger of the two, or some other figure? Some light on the answer may be shed by considering whether the occurrence of one is likely to be related to the other. If so, the surplus requirements would seem to be additive or at least greater than the larger; if not, then the larger of the two may suffice.

Ten years ago most actuaries, I think, felt that the use of surplus as a hedge against mortality fluctuation was really limited to the serious industrial accident and primarily within the group field. The flu epidemic, widely cited as a case in point, was, in the light of medical advances, generally viewed as of historical interest only and not as an example of what might recur in the future. In recent years, however, man's inability to control his own atmosphere is a cause for real concern. A great deal has been written but much less has been done about air pollution in our major cities. It is my personal conviction that what is now being done is far from adequate to control the levels of pollution and that a disaster causing a great many deaths in one of our major cities will be required to trigger the allocation of sufficient money and attention to cure the problem. Necessarily, it will be a number of years between the date of the first catastrophe and the date when sufficient attention has been given the problem to eliminate it. If this dire forecast is true, the life insurance industry must look forward to some fairly unhappy mortality experience at the catastrophic level from time to time over the next ten years or so.

In spite of feeling that mortality at a catastrophe level is a real pos-

sibility in the not-too-distant future, I still look upon surplus as primarily a hedge against asset depreciation for a company like the Phoenix Mutual with a preponderance of permanent individual insurance, a moderate amount of group, and a modest pension business. (Obviously, a company whose obligations are primarily short term, such as a pure accident and health insurer like the Blue Shields, would have an entirely different problem.)

Put another way, my feeling is that asset depreciation is the primary hazard for a company's permanent individual and pension business and would still be the primary surplus target until group life and health became a very significant portion of the company's total premium income. Further, it would not appear that substantial adverse mortality would be causally related to a severe asset strain, although some additional mortality would have to be anticipated. This leads me to the conclusion that a surplus fund adequate for severe economic depression should be satisfactory with minimal additions for other hazards. To be somewhat more specific, were it possible to simulate the effect of the Great Depression on our present portfolio, I would personally be happy with a current surplus position which would allow my company to weather such a catastrophe with only a minimum of surplus remaining, say, 1-2 per cent of assets.

W. HAROLD PHILLIPS:

I wish to compliment Mr. Trowbridge for his initiation of discussion in an area that requires a great deal of work. While the paper is termed "theory," in many respects it is a very practical approach to the surplus question.

Two different approaches to the question are referred to in the paper: "This contribution to surplus can be the result of a direct charge for surplus in the dividend calculation or it can come about indirectly, by using somewhat conservative claim, interest, and expense factors in the dividend formula." As our actuarial science develops and matures, and with the help of the computer, it would seem appropriate that we move from a "conservative assumption" approach to a "direct charge" approach. Using conservative assumptions, you never quite know where you stand except in an over-all sort of way. The direct-charge approach compels us to seek answers to the questions raised by the paper: How much surplus do you need and what for, where does it come from, how do various classes contribute to it equitably, who does it belong to, and how is it disposed of?

This discussion focuses attention on the direct-charge approach and some of its implications.

The Nature of Surplus: A Risk Premium or a Temporary Withholding?

Should contributions to surplus be considered in the nature of a risk premium and as such perhaps nonreturnable? Or is surplus something that is held temporarily while needed and later returnable on termination since it is no longer required? Are surplus contributions considered as belonging to the insuring organization for its corporate integrity as opposed to belonging to individual members or classes? How is all this related to the theory of terminal or termination dividends? The analysis below may assist in providing answers to these questions.

Let us work within the following framework, using a direct-charge approach and making the following assumptions:

1. The required level of surplus, the manner in which it will be accumulated, and a means of equitably charging surplus needs to cells and blocks have been developed and decided upon.
2. The various kinds of contingency needs are additive. The overconservative nature of pure addition alluded to in the paper has already been discounted by a lower level for some or all of the kinds of contributions.
3. Surplus is distributed by use of a fund technique: the fund (on target at the beginning of the year) together with interest provides death benefits, expenses, withdrawal values, and contingency contributions. The amount above that required to put the fund on target at the end of the year can be considered the dividend payable.
4. The asset-share accumulation includes contingency funds.

Within this framework we might have two categories of contingency contributions:

- A. Contributions that are a continuing charge payable by all survivors in a group. These can take the form of some or all of the following:
 1. Per M;
 2. Per contract;
 3. Per cent of premium;
 4. Points on the interest rate, e.g., if we assume an earned rate of 4.80 per cent and require 15 points as a contingency contribution, then 0.0015 of all assets would accumulate in a contingency fund at an interest rate of 4.80 per cent.
- B. Contributions that are required to maintain the contingency fund at a certain level for survivors. Amounts not required to maintain the contingency fund at the desired level just for survivors can be thrown off. These could take the form of some or all of the following:
 1. A per cent of reserves;
 2. A per cent of assets;
 3. A per cent of the amount at risk or a function of the amount at risk varying with age.

The answers to the questions posed earlier in this section are implicit in the methods or formulas used in the accumulation of the various kinds of contingency funds.

Type A contributions can be considered fully nonreturnable. The contributions by all survivors simply continue to accumulate; these are shown in the asset-share accumulation as an accumulated amount per 1,000 surviving. At longer durations, where the number of survivors out of 1,000 starting are quite small, the amounts accumulated can be quite large. The accumulator takes the form of $1/(1 - q - q^w)$, where q^w is the rate of withdrawal other than death (q). However, if the insurer believes in a philosophy of terminal dividends, even Type A accumulations could be returned to the terminators (surrender and/or death) all or in part. In this case, the formulas would take a different form. The terminal dividend would be withdrawn from the contingency fund, leaving a smaller fund for survivors. Note that if this approach is taken for Type A contributions, it may be necessary to set the original level of these contributions at a higher point than otherwise.

Type B contributions release the accumulations for those that terminate automatically, since they are maintained at a certain desired level only for survivors. The contingency funds released for terminators can be paid either to the terminators in the form of a termination dividend or can serve to increase the dividend payable to all. In both cases, the same amount is released but to different people.

The following position could be taken from the possibilities suggested above:

1. All contingency contributions are in the nature of a risk premium and belong to the insurer (as contrasted to individual members) in order that it be able to maintain its corporate integrity. Nevertheless, we will attempt to keep track of all contributions through the asset-share accumulations in order to allocate them equitably.

2. Type A contributions are nonreturnable. They lend themselves to providing the surplus necessary for new-business investment (discussed in the next section). When Type A contributions are used for the purpose of new-business financing, are they then a risk premium? This nonreturnability feature ties in with the ever expanding needs for surplus to finance new business in a growing organization.

3. Contingency funds accumulate for the benefit of survivors and future entrants. No terminal dividends are payable.

Financing New Business (Investment in New Business)

One of the most important aspects of surplus retention and accumulation is that required for new-business financing. I am not aware of its

being covered specifically and carefully in the actuarial literature. Jackson and Trowbridge allude to it a number of times but never come to grips with the problem. These remarks will attempt to fill in this gap. This section could be subtitled "Theory of New-Business Investment."

In order to assist in the discussion, let us consider Chart I, which can be used to represent a cell (plan/year/age), an entire year's business, or an entire block of business.

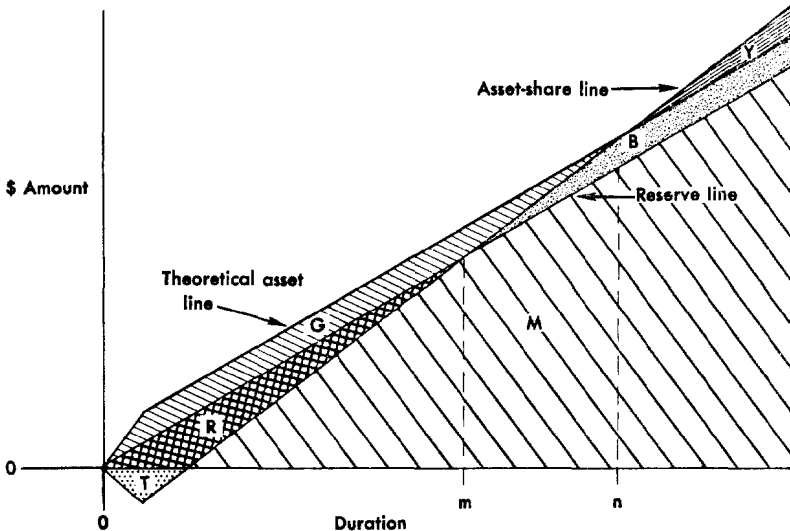
Definitions

Asset-share line.—This is the retrospective asset-share accumulation based on experience, or a projected retrospective asset share based on best-estimate assumptions, accumulating all contingency contributions. The sum of these lines for all cells and blocks equals the assets of the entire organization. These assets total all $M + B + Y - T$ areas. Thus $M + B + Y - T = \text{Assets} = A$.

Reserve line.—Legal liabilities (L). This item includes one-half of the following year's declared dividends. The NAIC statement is redundant to the extent of 50 per cent of the following year's dividends. Approximately 50 per cent of the dividends have already been earned and thus are proper liabilities. L excludes MSVR, which is treated as surplus (S). Legal liabilities total all the $M + R$ areas. Thus $M + R = L$.

Theoretical asset line.—These are the assets that theoretically should be held for a cell, block, and so forth. This equals all areas of $M + R + G + B$. Thus $A = M + R + G + B$.

CHART I



n.—The point in time where the asset share first reaches the point where the organization has accumulated the liability and necessary contingency requirements. This may be termed the “break-even point.”

m.—Point in time where the asset share first reaches the reserve.

Book loss.—Sum of all $R + T$ areas.

Investment in new business (or amount of financing of new business).—Sum of the $G + R + T$ areas.

Some very interesting relationships can be developed from this picture. By definition, assets are as follows: $A = M + R + B + G = M + B + Y - T$.

This is the ideal situation. All surplus distribution techniques and formulas work toward putting you in this position. Thus $R + G = Y - T$ or $Y = G + R + T$.

Now

$$\begin{aligned} A &= L + S ; \\ S &= A - L \\ &= M + B + Y - T - (M + R) \\ &= B + Y - R - T . \end{aligned}$$

Also

$$\begin{aligned} S &= M + R + B + G - (M + R) \\ &= B + G . \end{aligned}$$

The following comments result from this type of analysis:

1. The analysis assumes that the carefully constructed asset shares for all business in force total the annual-statement assets. This is an immense task.
2. In a growing organization, a significant portion of business in force is at durations under n . The business past duration n must have generated enough assets to cover the “deficits” for business under duration n .
3. Significant amounts of surplus are required to finance new business. Surplus is required not only for the kinds of contingencies discussed in the paper but also for new-business financing.
4. Surplus requirements for new-business financing can exceed requirements for other contingencies.
5. The investment in new business (strain on surplus) exists whether reserves are net level or CRVM. Of course, net level strain is much more severe.
6. Tradition and competition dictate that dividend distribution start at duration 1 or 2, while theoretically, perhaps, it should be withheld until n or m . This earlier distribution affects the time at which you reach n .
7. The target in the dividend-distribution process is to attain the proper contingency level together with required liabilities at duration n . What should

- n be? In practice it might be set at 10, 15, or 20. The point m has significance only in relation to book position.
8. The amount invested in new business continues to increase as long as growth continues. This requires ever increasing amounts of surplus, which are locked in and thus cannot be returned (as long as growth continues).
 9. The ever expanding investment in new business also occurs in a stock nonpar situation. An ever growing amount is locked into new-business financing. Will the stockholders ever have access to it unless the growth rate starts to decline?

Effect of Growth Rate on Competitive Position

I would like to emphasize one point made by Mr. Trowbridge. This section is a very practical application of the theory developed in the paper.

The variation in g can have a dramatic effect on competitive posture, as shown in the accompanying examples. As developed in the paper, the additional surplus contribution in total over and above that provided by interest on surplus itself is gfP . This can be developed further:

$$\begin{aligned}\Delta S - iS &= gfP - iS \\ &= gfP - ifP \\ &= fP(g - i).\end{aligned}$$

This defines the total contingency requirement in terms of a margin in the interest rate. If we assume that (a) all contingency requirements are measured by fP , (b) all contingency charges can be made through a margin in the interest rate, and (c) all loading and mortality gains are accounted for in the dividend distribution, then the insurer can assume an i' in the dividend formula of $i' = i - f(g - i)$.

EXAMPLES

	A	B
g	0.10	0.05
f06	.06
i	0.05	0.05

The size of P in the examples does not matter. A can pay an i' of 0.0470, while B can pay an i' of 0.05.

The paper contains the statement "Surplus thus has a role in a company's competitive position; by improving it once the surplus is largely built even as it hurts it during the period that surplus is being accumulated." I would suggest that as long as an organization continues to grow, surplus will never be "largely built" and "the period during which surplus is being accumulated" is forever. Whether the growth "hurts"

depends on whether $g > i$. The degree to which it hurts depends on how much g exceeds i .

CHARLES A. WELSH:

Mr. Trowbridge is to be commended for bringing to our attention the actuarial philosophy in surplus accounting. This topic is particularly timely in view of the current interest in corporate planning.

Historically, the primary concern of the regulatory authorities in the insurance industry has been to assure solvency. Consequently the reporting procedures prescribed for statement accounting tend to err in the direction of understatement of surplus, particularly where some doubt exists with regard to the true value or where the value of the item is thought to be subject to fluctuation. This concept of surplus is peculiar to the insurance industry and is generally regarded as being deficient as a measure of net worth. Nevertheless, for certain purposes a knowledge of the amount of statement surplus is essential to company management, since it represents the amount of available surplus.

On the other hand, for certain internal purposes (for example, the valuation of corporate worth accompanying liquidation proceedings) the use of statement surplus would be inappropriate, and it becomes necessary to know the true surplus on an accounting basis comparable with that used in other industries. The exact nature of the true corporate net worth is subject to differences of interpretation, but a partial list of the type of adjustments necessary to reconcile the true surplus and the statement surplus would include (1) inclusion of nonadmitted assets, (2) recognition of value of existing business in force, and (3) elimination of certain adventitious liabilities, notably the Mandatory Securities Valuation Reserve.

On a liquidating basis, an investment-fluctuation fund is clearly a part of corporate net worth. However, the amount of the Mandatory Securities Valuation Reserve is fixed by regulation. Therefore, for purposes of determining availability of surplus, it is just as real a liability as any of the other items above Line 26. This distinction between liability and surplus does have its practical side for mutual companies operating in New York, since as a liability these companies are able to maintain this fund without worrying about compliance with the requirements of Section 207 for distribution of surplus to policyholders. The answer to the question of whether the MSVR should be treated as surplus or liability really depends upon whether a statement-surplus or a true-surplus concept is intended. The implied question seems to be whether there should be a MSVR in the first place, and this remains a subject of controversy.

The proper amount for contingency funds is a subject about which

there is no universal agreement. From general reasoning it is obvious that the probability of financial ruin is inversely proportional to the amount of the contingency funds, but the degree of confidence which can be ascribed to any particular amount is not readily known since these funds are subject to other than statistical fluctuation. Accuracy in actuarial calculations is, of course, a relative thing. The valuation of policy reserves, for example, employs the most highly refined actuarial techniques available. Generally speaking, the smaller or more elusive items are more likely to get a lick and a promise. Of all the items in the balance sheet, the contingency funds are among the least amenable to actuarial techniques. Consequently most companies use some approximation such as X per cent of assets or Y per cent of premium for these items.

The trouble with approximate valuation techniques is they usually do not have any intrinsic validity as standards of value. Further, they tend to become ensconced in the ritual of insurance accounting and remain long after they should have been replaced. With the current capabilities of electronic equipment we are in a good position to start phasing out the rules of thumb and phasing in more scientific valuation procedures. I can see swinging along for awhile with \$65 per thousand, $3\frac{1}{2}$ times premium, and some of the other old standbys, but I cannot see borrowing new rules of thumb from outside the life insurance industry when we already have too many of our own. If the contingency reserves cannot be valued with existing actuarial or statistical theory, then model-office or Monte Carlo techniques should be used.

The idea of a policy-based contribution to surplus has its analogue in the contribution theory of dividend distribution, and the various pros and cons of the individual-policy approach have been admirably presented and discussed in Mr. Jackson's paper (*TSA*, Vol. XI). It may be of some interest to consider the policy-based contribution to surplus for ordinary insurance as being represented by the individual policy's impact on the surplus account through the operating gain. On a policy-year basis the emerging surplus for a whole life policy could be expressed as

$$\begin{aligned} G_{[x]}(1 - c_t) - e_t - vq_{[x]+t-1} - v\dot{p}_{[x]+t-1} {}_tV_{[x]} - {}_{t-1}V_x \\ = v\dot{p}_{[x]+t-1} {}_tS_x - {}_{t-1}S_x, \end{aligned}$$

where

- $G_{[x]}$ = Gross premium at issue age x for a whole life policy;
- c_t = Rate of commission and other percentage expenses in policy year t ;
- i = Earned interest rate and $v = (1 + i)^{-1}$;
- $q_{[x]+t-1}$ = Rate of mortality in policy year t for a life age $x + t - 1$;
- e_t = Expenses (as of beginning of year) in policy year t ;

${}_tV_{[x]}$ = Policy reserves, including miscellaneous reserves;

${}_tS_x$ = Policy's share of surplus, including contingency-funds profit, and divisible surplus as of beginning of policy year t .

From this admittedly simplified equation the progress of the surplus funds can be followed. Owing to the level provision in the premium for expenses which are heavily skewed toward the first policy year, a surplus deficit is incurred at issue for almost all plan-age combinations in ordinary insurance. The importance of the initial surplus drain is that the insurer's ability to grow is limited by the amount of free surplus available. In a stock company this initial surplus drain is in effect advanced by the stockholders. When the cash flow reverses, as is usually the case in the second and later years, the policy's contribution to surplus is allocated partially toward restoration of the initial surplus drain and partially toward profit. Profit in turn is either paid to the stockholders in dividends or remains as part of the corporate worth and is available for growth of the business.

In a mutual organization, if it is accepted that there are no profits, the source of growth funds is less apparent. In order to return all contributions to their source, the present value of dividends must equal the present value of contributions to surplus. By suitably choosing the incidence of dividends, however, the insurer is able to control the incidence of the funds available for procuring new business. Thus it is mathematically possible for an insurer to sustain long-term growth while at the same time adhering to the philosophy of returning eventually to each generation of policyholders their entire contribution to surplus. Historically it is doubtful whether this philosophy has been sedulously pursued in practice. Moreover, there is a degree of uncertainty surrounding the proper level of the free surplus and contingency funds, and in such situations insurers are wont to defer distribution until the extent of the risk is more exactly known. For these reasons I would be inclined to include in any theory of mutual surplus the recognition that a portion of the corporate surplus has not been contributed by the current generation of policyholders and is not returnable to them except in the event of liquidation.

At the risk of seeming captious, there is one minor point which I would mention. Mr. Trowbridge has stated, "To think of contingency reserves as multiple rather than singular seems unnecessarily confusing." It would seem to me that logically a method for determining the amount of surplus must begin with the identification of the various purposes for which the surplus is held. The mere fact that a part of surplus is earmarked for a specific purpose implies neither that the funds so designated may not be diverted to another purpose nor that management's control over these

funds is anything but plenary. I hope that the author will clarify this point.

Again I would like to commend Mr. Trowbridge for bringing this topic within the purview of actuarial literature. I enjoyed reading his paper.

ROBERT F. LINK:

It is always a great pleasure to study a paper by Mr. Trowbridge. He has a faculty for formulating the principles of a confusing subject. This discussion is intended to build a little on his foundations.

Mr. Trowbridge asks, "What is surplus?" He considers whether it should be measured by a conservative approach or a realistic approach and concludes that "the discussion of other surplus questions following does not hang on the particular definition of surplus chosen." Others would go further and say that the fundamental question is what assets are needed in total and that the separation into reserve and surplus is a quite secondary question. Such statements challenge traditional views and stimulate us to go back to first principles in examining the nature of surplus.

My examination suggests that there are three distinct layers or parts of surplus. I will amplify this statement, but first some definitions may be useful.

Consider the assets, reserve, and surplus of a hypothetical insurance company, B , where A is the amount of assets; R_G is the reserve on Mr. Trowbridge's "best-estimate assumptions and prospective gross-premium-valuation methods"; $S_G = A - R_G$ is the surplus on the G basis; $R_M > R_G$ is a reserve with margins, particularly the reserve held for annual-statement purposes; and $S_M = A - R_M$ is the surplus on the M basis.

Let $P(x)$ be a continuous density function representing the probability that assets now on hand of amount x , together with future investment return and future annual premiums will be exactly sufficient to pay benefits now contracted for, plus related expenses.

The probability recognizes not only chance variations but also the likelihood of alternative future environments as to interest, experience levels, and so forth, and is on a "true," i.e., marginless basis. Then

$$R_G = \int_0^{\infty} xP(x) dx.$$

Let $PS(A)$ be the probability that assets now on hand of amount A will be *at least* sufficient. Then

$$PS(A) = \int^A P(x) dx.$$

Surplus: Part I

S_G is the net worth of B on a balance-sheet basis, the mathematical expectation of ultimate gain or loss. Its significance is probably clearer for a nonparticipating company than for a participating company. The same sort of test can be made for a block of business or even one contract, using as A the asset share or fund account. In this case it indicates whether fund-account transfers should be considered.

The G basis leads to another statement. "If $S_G = 0$ (i.e., $A = R_G$), then $PS(A)$ is of the order of $\frac{1}{2}$." We cannot say that it is $\frac{1}{2}$ because $P(x)$ is presumably skew. An actuary whose company valuation revealed a probability of $\frac{1}{2}$ that assets are sufficient might feel like a sky-diver who, while in free fall, learns that there is half a chance that his parachute will work.

This example clearly reveals the need for surplus relative to a reserve R_G . This surplus should be sufficient so that $PS(A)$ is very nearly unity. How do we determine the amount of this "sufficient" surplus?

As Mr. Trowbridge's analysis suggests, any amount held because of uncertainty as to the level of interest or mortality (morbidity) must recognize by class of risk the duration of exposure to risk, the risk insured against, and the amount at risk. Extremes are term insurance and paid-up deferred annuities. There are all shades of variation between. The necessary surplus level at this stage very nearly varies by valuation cell. We therefore get our major "surplus" by computing a reserve with interest and mortality margin.

Inclusion of margins yields an R_M , as defined above. This R_M will normally be the annual-statement reserve. Though the actuary may not compute his $PS(R_M)$, his choice of assumptions assures him that it is large enough. In other words, he feels that his company is safe enough if $A = R_M$ (or $S_M = 0$).

Surplus: Part II

It becomes immediately apparent that we need some *more* surplus. If S_M is zero, any tremor of experience could turn it negative. By convention, incorporated in practice and law, a negative S_M is unsatisfactory. Thus, the *second* purpose of surplus—in this case S_M —is to reduce the probability that short-term events will cause A to become less than R_M . This encompasses the familiar purposes—to protect against asset losses and catastrophic experience, to facilitate expansion of operations, and to provide a base for risk-taking.

I think that this second layer of surplus is the true domain of Mr. Trowbridge's formula

$$S = \sum_1^n f_k P_k .$$

Mr. Trowbridge has suggested assets and premiums as possible parameters P_k . Additional possibilities are amount insured, amount at risk, and expenses. Mr. Trowbridge suggests that the f_k relating to assets may differ by asset classification. Another area of possible differentiation relates to the risk of a calamitous mortality fluctuation. This risk seems much smaller on annuities than on insurance, because (1) not many calamities cause a sudden drop in mortality and (2) in any event, the mortality ratio cannot drop below zero. There could be a serious long-range drop in mortality, but this contingency may be recognized in reserves.

Surplus: Part III

A company's dividend formula establishes asset shares, the sum of which defines the total assets of the company (with certain reservations, depending on what is included or excluded in the sum). The dividend formula must be such as to result in a sufficient S_M . If the formula is such as to create relatively large asset shares, S_M may be greater than strictly necessary by the Part II criteria. The excess is Part III, the portion of surplus held as an adjunct of dividend policy. As Mr. Trowbridge noted, one value of additional surplus is to earn additional interest that may be used to pay higher dividends.

Summary of Parts of Surplus

The discussion thus far may now be summarized. Surplus relative to a reserve on a most-probable gross-premium basis may be regarded in three parts.

Part I serves to reduce the long-term risk of insufficient funds to an acceptable level. This part is usually incorporated in reserves computed on assumptions with margin. It should be so incorporated, in order to have a signal of any impending danger.

Part II protects against short-term adverse experience trends and hopefully prevents assets from falling below the desirable level established in Part I. The desired amount of Part II may well be determined by Mr. Trowbridge's basic linear compound formula.

A Part III may be held, reflecting the actuary's choices of rates, dividend formula, and reserve basis. This part probably is not set as an

explicit objective but results from the interplay of the various relevant factors.

Application to a Line of Business

The foregoing is theoretical. Can it supply any practical conclusions applicable to actual operations? I tried it on group annuities, with modest results. For simplicity, I assumed a company that has only group annuity business of a traditional character, that is, deferred annuities and deposit administration contracts with fixed guarantees (no IPG, separate accounts, etc.).

It may help to summarize first some special attributes of group annuity business. (1) It is single-premium annuity business, involving liabilities of great duration. (2) It is not fully "structured." An individual policy always belongs in one defined cell. In a group annuity contract, considerations, reserves, and benefits depend upon emerging events. The variation in size, age, form of benefits, and other factors is extreme. (3) Rates may change from time to time for future considerations. The rate-basis mix of each contract is the unique result of the historical factors of timing and growth applicable to the case. Furthermore, approximate rates are often used. (4) Discontinuance is interesting in that (a) liabilities usually run on for many years and (b) you do not always know whether a contract has discontinued (particularly if it is deposit administration). Note also that some contracts make large sums subject to cash withdrawal, on varying terms of approval or spreading. (5) Group annuity dividends are usually determined by a special form of dividend formula, described briefly later in the paper.

One suspects that Part I surplus ($R_M - R_G$) should be a relatively large proportion of R_M . There are two reasons: (1) the effect of margins increases with the duration of liabilities and (2) mortality trends, being unfavorable from the liability viewpoint, must be recognized by a projection that may turn out to be redundant. This could be verified or disproved by a suitable model-office test.

Part II should probably be relatively small. As pointed out earlier, the danger of short-term catastrophic mortality is limited for annuities by the nature of the possible catastrophe and by the absolute limit of expected mortality.

Part III is a function of the special dividend characteristics of group annuity business. These characteristics may be fairly well illustrated by the dividend formula of my company.

An asset share or *experience fund* is maintained for each contract. Actual receipts and disbursements under the contract are recognized, sub-

ject to (1) appropriate allocation procedures for investment results (the investment-year method) and expenses and (2) an adjustment to compensate on a credibility basis for a portion of mortality gains and losses. A reserve for future benefits, expenses, and contingencies is held against the experience fund. This reserve is on a special basis that recognizes the applicable characteristics of the case, including size parameters and the investment-year composition of the experience fund. Any excess of the experience fund over the reserve at the end of a calendar year forms the basis of the dividend on the case.

This dividend approach makes every contract a block of business B for surplus measurement. Our objective is to maintain the experience fund A at a level such that $PS(A)$ is high for each contract. However, any reasonable A will present some slight prospect of ultimate loss. This probability will vary from case to case, reflecting size, age, and other characteristics. The present value of ultimate losses on a case (mathematical expectation basis) is

$$PVL(A) = \int_A^{\infty} (x - a)P(x) dx.$$

There is a complementary expression for gains which, under this approach, generally flow into dividends in due course.

The risk of loss must be paid for. The deficits on contracts that terminate in the red must be balanced by transfers from the experience funds of other contracts. Such transfers can take the form of small annual risk charges, which can be accumulated in a special risk charge fund. The special risk charge fund should be large enough so that, with anticipated future additions thereto, it will cover anticipated losses and any other applicable contingencies. Thus, the A for our group annuity company (subject to bookkeeping adjustments) is equal to the sum of the A 's (experience funds) for contracts plus the special risk charge fund.

Would our group annuity company have a Part III surplus, or do overall company needs control the dividend formula? Probably the answer is a mixture in most companies. (The annual statement is a hard master.) Here are a few factors, with no opinion on how they should add up:

1. $PVL(A)$ as a per cent of A for the company will probably be much less than the average per cent for contracts (a) because risks are combined in the calculation at the company level and (b) because we have the special risk charge fund in addition to experience funds. This, of course, is as it should be. A much higher level of risk can be accepted at the contract level than at the company level.

2. For an average distribution of business, the company probably does not have the full experience funds called for by its dividend-formula objectives.

3. Companies will differ in their degree of risk-taking at the contract level. A company with relatively low dividend-reserve objectives and high $PVL(A)$'s (and, one trusts, higher risk charges) is more likely to find its over-all company financial needs controlling.

By way of summary, group annuities would probably require a relatively large Part I surplus and a relatively small Part II surplus. The Part III will depend on the collection of factors discussed above.

Finally, one can heartily second Mr. Trowbridge's suggestion that there is a large gap between this embryonic theory and the actual practice. The theory will be barren unless it develops sufficiently to assist in the practical decisions that must be made annually in the conduct of company business.

ABRAHAM HAZELCORN:

Mr. Trowbridge's surplus theory will be of use as a guidepost to actuaries confronted with this problem. While the paper is devoted to a theoretical area, it is well to note that in the matter of returnable versus nonreturnable surplus, practice has, in a sense, led theory at times. This was highlighted by the presentations made in regard to terminal dividends at the New York Insurance Department about ten years ago.

The diversity of dividend-distribution methods did not remain entirely within the framework of practical solutions to an over-all unified theory. On the contrary, testimony, in my recollection, indicated that the basic concepts of surplus and whether or not a policyholder is to make a permanent contribution to surplus found different answers among some of the leading companies.

Perhaps this is as it should be. The basic management decision after provision for what is considered in each company actuary's mind the absolutely necessary contingency reserve should be available for maintaining or increasing the company's vitality. This, of course, involves the concepts of equity. Some actuaries within their concept of broad equity will find enough room to operate; other actuaries will deem that some solutions violate basic equity. The reputation of the company in the form of its dividend illustrations versus history will be a considerable force here also.

Perhaps it is with this attitude that the New York statute does not usurp the management decision of the amount of returnable surplus. It, however, examines the dividend methods used to achieve equity once that amount is decided upon. While there is a limitation on the accumulation of surplus at 10 per cent of policy reserves and policy liabilities in Section 207 and the statute (Sec. 216) concerning policyholders participa-

tion and surplus in life insurance companies, there is no direct regulation of "such sums as may be deemed advisable for the accumulation of a surplus not in excess of the maximum prescribed in this chapter. . . ."

At least one large life insurance company maintained a constant relationship between surplus and liabilities. This was according to the NAIC statement. It may well have been redundant since, as one examiner put it, "The liabilities were dripping with assets." Therefore, Mr. Trowbridge's point about a more realistic valuation of assets and liabilities is well taken.

I do not agree with Mr. Trowbridge's statement concerning the preservation of equity between succeeding generations of stockholders in a profit-making corporation. While there is a marketable ownership share, timing can probably cause greater dislocations of equity than in a mutual life insurance company. The vagaries of the market plus a lower percentage of earnings payout based on a greater image of growth will, I feel, do more violence to most concepts of equity than the diversity of life insurance dividend philosophy.

BERT A. WINTER:

Those of us employed by mutual life insurance companies are much indebted to Mr. Trowbridge for his penetrating discussion of a subject so central to our professional responsibilities but on which the published literature is, as he says, relatively silent. He has lucidly organized the subject into five questions and, under each, has quite properly, in terms of his stated purpose of a general development, given the full range of considerations that might influence management in arriving at the current answer for their company to that question. The purpose of this discussion is to indicate circumstances, primarily with respect to individual life insurance policies, influencing the choice of options within this range.

What Is Surplus?

I should like to define the minimum reserve for a company's policies providing contractually guaranteed surrender values as:

The 1958 CSO $3\frac{1}{2}$ per cent CMVR in the first policy year, with a modified valuation premium thereafter that will accumulate to the guaranteed cash-surrender value at duration c , the earliest policy duration at which the cash value exceeds the CMVR on this mortality and interest basis. After duration c , the mortality and interest basis is that underlying the guaranteed cash-surrender value at the longest premium-paying durations, and the valuation premium the amount that will accumulate to the net level premium reserve on this basis by the end of the premium-payment period, or earlier duration at which the cash value becomes equal to the net level reserve.

I believe that the reserve on this basis is the lowest that could properly be held after the occurrence of a catastrophe for the policies then remaining in force. Hence, I think it illusory for management to rely on any greater surplus than that derived from such a reserve, as provision for possible future catastrophes, however probable the greater future profits anticipated. This is because it is unlikely that the company could survive as a continuing organization to enjoy such then future profits, after the occurrence of a catastrophe severe enough to reduce its accumulated funds below the amount required to hold the reserve on this basis for the remaining policies.

Accumulated Issue Strain

The excess of reserves actually held over the minimum reserves just described may be regarded as the "net accumulated issue strain." It varies primarily with the relationship of new business to in-force business and hence is managed primarily by changes in the level of early dividends as compared with "ultimate dividends" and changes in the slope by policy duration with which the former approach the latter rather than by changes in the level of the ultimate dividends themselves.

Ultimate Surplus

Thus "ultimate surplus" may be regarded as the sum of assigned and unassigned surplus in the published statement, investment valuation reserves held as liabilities or deducted from assets in the published statement, and accumulated issue strain.

This ultimate surplus may, as Mr. Trowbridge says, be expressed as the sum of an asset-related term and a risk-premium-related term. However, for purposes of assigning it to individual dividend cells, both as a fund-accumulation objective before "ultimate dividends" are granted and as a measurement of the funds assumed present in the calculation of ultimate dividends, I have found it convenient to change the form slightly. Thus, the base of the asset-related term may be taken as net level premium reserves, as more readily susceptible of calculation for individual cells. The base of the risk-premium element may be taken for permanent policies as the tabular cost for the policy year just ended, which gives a better distribution over the premium-paying and fully paid periods of limited payment policies and measures more appropriately the disappearance of the risk element as endowment maturity or the a -point of a retirement income policy is approached.

ALLAN F. LEBOURVEAU:

Mr. Trowbridge is to be congratulated for his thought-provoking paper. In response to his invitation to add to the concepts he has expressed, I have outlined an approach that has proved both practical and useful. To simplify the outline, I have considered only individual life insurance; any modifications needed for group and health insurance will generally be apparent.

Basic Considerations

All operations of a mutual life insurance company are designed to carry forward its relationship to its policyholders. The assets of any company will equal the excess of income over disbursements over the whole period during which the company has existed. For a mutual life insurance company, this means that the assets available for its use are equal to all the premiums received from policyholders less benefit payments and dividends to policyholders less the cost of services rendered to and for policyholders (that is, expenses and taxes) plus investment earnings on the resultant accumulations.

Why were the assets of a mutual life insurance company accumulated? They were accumulated so that the company can meet its obligations to its policyholders. This is a complete answer—within the limits prescribed by law, a mutual life insurance company is justified in holding assets in the amount that it feels necessary to carry out its obligations to its policyholders, but anything more or less than this should be paid out or recovered through adjustments in the amount paid out in dividends.

But if this broad concept is a proper one, how do reserves fit into the picture? Statutory reserves are an aggregate evaluation of the guaranteed benefits in all the company's policies according to definitions set forth in laws primarily intended to establish criteria for company solvency. In the aggregate, statutory reserves constitute the largest part of the amount which the company, as a going concern, considers is needed to meet its obligations to its policyholders, but, as will be shown later, statutory reserves are a very crude and imperfect guide to the amounts that should be held by a company to meet its obligations on each class of policies.

What is surplus? Arithmetically, the annual statement would indicate that it is equal to the total assets less the total reserves; that is, surplus is equal to the total amount held by the company to carry out its obligations on each and every class of policies less the total of amounts (the statutory reserves) which do not closely reflect these obligations for each class of policies. When surplus is expressed in this manner, it becomes apparent why it is difficult to develop a theory of surplus under which the

total amount needed for each class of policies is not considered, but only this balancing item.

I think, however, that the question of what surplus is in a mutual life insurance company should really be considered much more deeply. Everyone knows the meaning and purpose of capital stock and surplus in the usual corporation which has stockholders. But a mutual life insurance company has no owners; its policyholders have only such rights as may be given them in the insurance contract and by statute. Accordingly, a mutual insurance company has no right to accumulate funds to be "owned" by its policyholders, but only such amounts as it feels that it needs to carry out its obligations to these policyholders. In other words, all of a mutual life insurance company's assets are needed to furnish insurance to its policyholders; thus, any excess of assets over the reserves which may appear in the annual statement is in no way surplus in the sense used in a corporation with stockholders.

*Obligations of a Mutual Life Insurance Company to
Its Policyholders*

The minimum amount that any life insurance company, stock or mutual, should hold would be the amount it considered necessary to cover the guaranteed benefits payable under its policies. In this sense, it can be said that the primary role of surplus is solvency, both present solvency and solvency in the future.

I think that the obligations of a mutual company to its policyholders go considerably beyond solvency. Participating premiums contain margins which would normally minimize this risk of insolvency. In consideration of these premiums, the participating policy not only contracts to pay the guaranteed benefits but also "to ascertain and apportion any divisible surplus accruing on this policy" and to pay it as dividends. While this does not create an obligation to pay a dividend every year on every policy, I believe that it does obligate us to operate to the best of our ability toward the goal that each and every class of policies be self-supporting and receive insurance at a cost that reflects its fair share of investment earnings, claim costs, expenses, cost of catastrophes, and so forth. When a mutual life insurance company accumulates funds with this additional obligation in mind, it will normally have funds enough, but no more than sufficient, to carry out all its obligations to its policyholders.

How Funds are Accumulated

As previously mentioned, the only way that a mutual life insurance company can accumulate funds is from contributions of its policyholders.

Underlying the traditional three-factor dividend formula, for example, is the assumption that the policyholder contributes the tabular net premium, thus enabling the company to accumulate the statutory reserve. In practice, the policyholder may contribute more or less than the tabular net premium as a result of differences between the various factors of the dividend formula and actual experience.

For many years my company has been using a modification of the contribution formula that reflects the fact that the statutory reserve is neither the amount that has been accumulated nor the amount that should be accumulated. In the early policy years, the actual amount accumulated is less than the reserve because of the higher level of expenses, particularly in the first year. At the longer durations, the reserve is only enough to cover the cash value, while the minimum amount that should be accumulated would certainly cover both the cash value and the expense of terminating the policy. The actual amount that should be accumulated on a class of policies would depend on the company's judgment as to the amount needed to carry out its obligations to each particular class of policies. Use of such amounts is the heart of the method that we use to accumulate funds—through policyholder contributions made as a result of using what we call "dividend funds" in our annual-dividend formula. The amounts accumulated in this manner are sufficient in the aggregate to cover the statutory reserves shown in our annual statement and the larger part of the amount shown as unassigned surplus.

In our annual-dividend formula we also make specific charges for catastrophes—events that may or may not happen during any particular period of time but for which we think a small charge in the nature of a risk premium should be made. We have made specific charges for such catastrophes in our dividend formulas for about twenty years, and the funds accumulated are sufficient in the aggregate to cover the amounts of investment-contingency reserves shown in our annual statement plus a part of the amount shown as unassigned surplus.

The Dividend-Fund System

The dividend-fund system is designed to reflect in a practical way the contributions of the various dividend classes to the surplus earnings and accumulated surplus of the company. Annual-dividend scales and terminal-dividend scales are both integrated parts of this system; they are based upon and reflect the same dividend funds.

Annual dividends are determined as the excess of premiums plus interest on the dividend fund over mortality and expense charges, a charge for catastrophes, and the planned increase in the dividend fund. Terminal

dividends generally reflect the excess of the dividend fund over the guaranteed cash value and charges for expenses and antiselection at termination.

Just as annual dividends can be considered the adjustment of the contractual premium to the premium actually needed, so terminal dividends in the dividend-fund system can be considered as the adjustment of the contractual cash value to the cash value actually available as reflected in the dividend fund.

Determining the Dividend Funds

The cornerstone of this dividend-fund system is the accumulation of funds according to a predetermined plan for each dividend class—that basic group of policies for which dividends are identical because all essential characteristics (policy series, plan, age, year of issue, etc.) are identical.

The predetermined funds should reflect all the factors believed to have a bearing on future needs. Among the factors to be considered are the long-term trend and possible range of variation in interest rates and mortality experience, potential extra costs of the exercise of settlement options, the possibility of increases in price and wage levels, as well as considerations of a more general nature. The accumulation of adequate funds is particularly important on policies which provide benefits beyond the end of the premium-paying period, when premium margins are not available to meet possible unfavorable experience.

If predetermined funds are to be realistic, they must also take into account the incidence of expenses. This means that in the early policy years the fund should reflect the fact that initial expenses reduce the amount that would otherwise be available. One overriding goal is that all dividend classes should be self-supporting even during a prolonged period of unfavorable experience.

After consideration of all the factors, the fund can usually be expressed in a relatively simple formula. For example, after the effect of initial expenses has been liquidated, the dividend fund might be expressed as the sum of (1) the reserve on the basis set forth in the policy, (2) a percentage of the reserve, and (3) a constant per thousand of insurance.

It cannot be emphasized too strongly that the payment or nonpayment of terminal dividends is not a consideration in determining these dividend funds; the only consideration is that the amount be that which in our judgment the company needs to insure the fulfillment of its obligations as previously described.

Catastrophe Reserves

As mentioned earlier, the dividend-fund system also makes provision for a catastrophe reserve. This catastrophe reserve is accumulated from specific charges (for example, a few basis points of interest plus a few cents per thousand of insurance) that are part of the annual-dividend formula. Capital gains are logically credited to this catastrophe reserve, while capital losses and war deaths are examples of costs that are logically charged against it.

A Comparison of Theories

At first glance, the dividend-fund approach that I have outlined bears a strong resemblance to the "policy-based" approach set forth by Mr. Trowbridge. This is, however, more a matter of appearance than a matter of fact.

While the dividend-fund system has predetermined funds for each dividend class and also makes provision for a catastrophe reserve, it has no fixed, over-all surplus goal for the company as a whole. As a result, the company's growth rate has no effect on its dividend scales. Under the theory of a fixed aggregate surplus goal, it is difficult to see how the principle laid down by Mr. Weeks so many years ago in Volume IX of *TASA* can prevail:

If a given company experiences year after year uniform rates of initial expense, renewal expense, mortality, and interest, then the dividends ought not to rise or fall according as the volume of new business is less or greater in one year than in another year.

A second major difference is in the purpose of surplus. Mr. Trowbridge views surplus primarily as an aggregate contingency reserve to enable the company to meet its obligations. Under the dividend-fund system, the first objective is to ensure that each dividend class will be self-supporting and will receive insurance at its fair share of the cost, that dividend distribution will be according to policyholder contribution; the objective of solvency is then met as a natural consequence.

Another significant difference is in the concept of "on target." Under the dividend-fund system, all policies are always "on target." If it should be decided that a strengthening (or destrengthening) of the predetermined dividend funds should be made, new levels of dividend funds could be established, together with a predetermined program of reaching these new levels. The dividend formula would then automatically and equitably adjust annual and terminal dividends to reflect the revised dividend funds.

Surplus Theory and Equity

The dividend-fund system provides direct answers to the questions of how, when, and to what extent each policy should contribute toward surplus. The dividend funds are the vehicle through which the company lives up to its obligation to apportion distributable surplus equitably.

Mr. Trowbridge points out three difficulties in the "policy-based" approach that might also be considered as applicable to the dividend-fund approach that I have outlined. First, he suggests that the concept of "individual returnable surplus" is in many ways foreign to the co-operative principle upon which mutual insurance is built. This is not, however, the concept of the dividend-fund approach, which is "distribution according to contribution." No individual accounts are maintained with individual charges and credits resulting in an individual balance to be returned. The dividend-fund system recognizes that, if each dividend class is to bear its fair share of all costs, it must contribute to surplus, so that the company will have funds adequate to meet its obligations. When policies terminate, the company's obligations are reduced and surplus may become available for distribution. Under the dividend-fund system the amount available for distribution at termination is determinable and is the basis for any terminal dividend payable. In my mind, there is no material difference in principle between a surrender dividend and a surrender value; while one may be considered as paid out of accumulated surplus and the other out of accumulated reserves, they are both, in logic, paid out because funds have been accumulated which are no longer needed.

The second difficulty suggested by Mr. Trowbridge is that dividends otherwise payable in the early years are hit hard by the necessity of rapid surplus accumulation; this is, of course, tied in with his concept of a surplus target expressed in simple rates without allowance for early expenses. Because the dividend-fund system takes these expenses into account and because the predetermined funds are, as a matter of policy, established to at least cover the cash value and the expenses of surrender, annual dividends at the shorter durations are, in fact, likely to be lower than under a dividend system that does not fully allow for early expenses. While this admittedly has its competitive disadvantages, it should be offset by higher annual dividends at the longer durations, if we assume that the over-all amount of surplus earnings available for distribution is the same.

Mr. Trowbridge also cites the difficulty in changing from another dividend system to a policy-based system. This difficulty, of course, exists in any change of dividend system, and such a change would only be rec-

commended by an actuary if he felt that the new system had advantages to the policyholder or company over the old system. While the dividend-fund system in its present form has been used in my company for only about twenty years, the theory of surplus on which it is based—that each dividend class should contribute sufficiently to be self-supporting—has been followed at least since the company was mutualized.

While this no doubt reduced our problems, the practical difficulties of adopting a dividend-fund system are not really very great when compared with the advantages the system has of insuring “distribution according to contribution” under both favorable and unfavorable economic conditions and with dividend classes varying radically as to premium level, reserve basis, and so forth. An actuary deciding to adopt a dividend-fund system would, no doubt, start by developing it for the series of policies currently being issued, where existing surplus would not be a consideration and where the problem of consistency with previous dividend payments would not exist. Once the system was developed for currently issued policies, the desired level of funds for earlier series could readily be determined and the problem would largely become one of developing a program for reaching this level of funds with a minimum impact on the dividends and net costs under the former dividend system.

J. STANLEY HILL:

Mr. Trowbridge is to be congratulated for his courage and intellectual vigor in tackling this most important and difficult subject. Although he emphasizes theory, his practical knowledge is much in evidence. Although he deals with retained surplus, the extensive treatment of the contribution to surplus has significant inferences with reference to surplus distribution. In short, he demonstrates that the two subjects cannot be divorced.

In answering the question “What is the purpose of surplus?” the author is silent on two important aspects: (1) the need for legal solvency and (2) the competitive need for an image of financial strength. These considerations may be classified as practical, but they have important influence on discussions affecting surplus policy—the former particularly for new companies and the latter particularly for those well established.

Mr. Trowbridge is extremely logical; his almost hypnotic approach should not blind us to other acceptable alternative approaches.

1. Instead of treating interest on surplus as a deduction from the needed contribution to surplus, a portion of the interest—or all of it—might be used to defray the cost of agency expansion. This avoids the less satisfactory and perhaps less justifiable method of charging this entire cost against current-year production.

2. The author's approach seems to imply that (at least so far as interest earnings are concerned), having determined the appropriate contribution to surplus, we can then solve for the amount to be distributed. No doubt he would emphasize that *in practice* one eye must be kept on the company's competitive position. Is there room for elevating to the theoretical level the philosophy that a mutual company can and should determine its pricing policy by first establishing its policyholder costs at a level which produces a predetermined relationship to those of the best-known companies and then solving for the amount of money available for growth? At first glance, this approach may seem to contravene the principle of mutuality. Further thought may lead one to conclude that it serves the interests of the present policyholders better than the more traditional procedures.

LOWELL M. DORN:

Mr. Trowbridge is to be commended for presenting an important paper. However, we are seriously concerned that his decision to make no reference to "court decisions affecting surplus accumulation and distribution"—and, in fact, to present certain theoretical concepts and inferences which are incompatible with well-established legal and actuarial concepts and practices in regard to surplus distribution—could hurt life insurance by serving to undermine those established concepts and practices in case of court tests.

As is well known, there is a long line of legal decisions, including *Rhine v. New York Life*, wherein the courts have laid down a definition of "classes" for the determination of equity in the matter of policyowner dividends. The legal definition of "classes" grew out of actuarial concepts and practices developed because of the necessity for mutual life insurance companies, among others, to treat their policyowners with equity as well as to meet the statutory proscription against unfair discrimination among policyowners of the "same class or equal expectation of life."

The legal definition includes within a dividend "class" all similar policies issued at the same time (in practice, usually the same calendar year) and under the same conditions, at the same age, providing the same benefits, and at the same annual premium rate per \$1,000. This accords with the concepts behind the "contribution method" of ascertaining and apportioning divisible surplus, with dividends per \$1,000 which vary by plan (including supplementary benefits and special restrictions), age, duration (or year of issue), and premium, reserve, nonforfeiture and underwriting classification.

This concept visualizes that each "class" receives dividends not in

accord with its own limited experience but based on the insurer's average experience with regard to interest, mortality, expenses, lapses, and so forth. Hence, each "class" helps support the over-all experience and is also supported by that experience. Moreover, termination dividends may be allowed, where indicated, without any requirement to pay out all of a terminating policy's share of the accumulated surplus in the "class."

For practical reasons, and within the bounds of equity, certain similar "classes" may be considered together and receive the same dividends per \$1,000.

This established concept of "class" is similar in many respects to the segment-based concept mentioned by Mr. Trowbridge but may also differ from it. However, it would exclude the policy-based concept and the company-based concept, except to the extent that a single policy, because of special conditions, is in a unique "class" all to itself for dividend purposes.

With regard to the section of the paper headed "Sources of Surplus," some actuaries might prefer to analyze contributions to surplus in a somewhat different manner than that described in the paper. For example, instead of treating interest on accumulated surplus separately, such interest might be combined with other investment income for dividend purposes. Moreover, contributions to surplus might arise from gains that cannot be described as either "a direct charge for surplus in the dividend calculation" or as something which comes about "by using somewhat conservative claim, interest, and expense factors in the dividend formula." Familiar asset-share calculations are designed, within practical limits, to reflect the various sources of surplus.

The paper also states that, under the policy-based concept, "the long-range surplus charge for any policy is *nil*." This theoretical statement could create serious misunderstandings in practice and mistaken concepts of equity. The fact is that no "class" of policies should expect to get back, through dividends, all its remaining contributions to surplus by the time the last policy in the "class" terminates. Rather, every "class" should properly leave a contribution to the surplus of the insurer, as part of the cost of insurance in a continuing enterprise.

GARNETT E. CANNON:

In his paper Mr. Trowbridge says that surplus has a role in a company's competitive position. In view of the federal income tax, which for most mutual companies results in a tax on investment income, it is worth examining this role.

It is our practice each year to make a determination of the rate of

interest for each classification of liability according to the manner in which it is affected by federal income tax. For example, last year our net interest rate was 5.60 per cent before tax and 5.02 per cent after tax. However, the aftertax rate return on life insurance reserves was 5.37 per cent, whereas it was only 2.97 per cent on surplus funds.

In computing asset shares, therefore, an interest rate of 5.37 per cent could be used on the portion of the asset share that is life insurance reserve and only 2.97 per cent on the portion that is represented by surplus. This has an important influence in determining the basis of distribution of surplus to policyowners.

Our studies show that in our case the net cost to the policyowner is lower if all dividends are paid annually than if a settlement dividend is paid. This assumes surrender of the policy at the end of twenty years with exactly the same balance remaining in company surplus under each method of distribution.

In other words, surplus held in our company that is not needed for growth and contingencies is heavily taxed. Therefore, we are looking for ways to get these earnings to the policyowners with as little dilution as possible.

(AUTHOR'S REVIEW OF DISCUSSION)

CHARLES L. TROWBRIDGE :

A review of the ten discussions attracted by this paper convinces the author of the validity of two of his basic concepts: first, the subject is an important and challenging one of real interest to actuaries, and, second, the subject is as "elusive . . . as it is fascinating" (Mr. Jackson's terminology) and one that can be expected to provoke contrasting points of view. Perhaps it will be helpful to review the discussion in terms of the five major questions proposed; in doing so, we may see to what extent the discussers appear to agree with the author—and with each other.

What Is Surplus?

The traditional view that surplus is represented by the excess of statement assets over statement liabilities gets a certain amount of support from Mr. Phillips and at one point from Mr. Welsh; but there is also support (from Messrs. Welsh, Link, Hazeltorn, and Winter) for viewing liabilities more realistically to produce a sounder base for surplus theory. The author has a personal preference, not particularly indicated in the paper, for the latter approach.

Mr. Winter's discussion implies that "accumulated issue strain" is really a part of surplus, or at least a part of what he calls "ultimate

surplus." Mr. Welsh speaks of an adjustment for "value of existing business in force." It clearly simplifies surplus thinking if the negative element to surplus arising from initial expenses can be eliminated—as it is if gross-premium-valuation methods are used in an internal surplus analysis.

Mr. Lebourveau and, to an extent, Mr. Link tend to de-emphasize the split between liabilities and surplus, or at least to emphasize the importance of equating their sum (which is equivalent to assets) to the amount needed to meet company obligations under adverse conditions. The author has no quarrel with this over-all concept and indeed looks at the problem this way himself. He finds it difficult, however, to judge whether the over-all concept can be considered adequate without some idea of a realistic liability under best-estimate conditions.

Mr. Welsh seems to prefer the practice of considering separately several pieces of what the author views as an essentially inseparable contingency reserve. Whether there is any real meaning in the allocations of surplus to certain specific (or unspecific) contingencies, as is common in life company statements and actually required by some states, we must leave to each actuary to determine for himself. The author views many of these as window dressing which may have some place in a statement prepared for state regulatory authorities and the general public but which should not mislead the actuary in his own analysis.

Purposes of Surplus

Mr. Hill adds the idea that surplus is needed for two reasons—to satisfy the test of legal solvency and to create an image of financial strength. For both these purposes it is clear that the statement surplus is the pertinent one. I assume that Mr. Hill would agree that any sensible surplus built along contingency-reserve lines is likely to meet these needs as it meets the others.

Mr. Phillips appears a little uncomfortable with my statement (in connection with the competitive aspects of surplus) that eventually the presence of a surplus helps a company's competitive position. He quite rightly emphasizes that the period during which surplus is being accumulated may be forever—and there is a net charge to policyholders so long as the growth rate g exceeds the interest rate i . If Mr. Phillips and I have any real differences, which I doubt, it lies in my inclination not to assume perpetual growth, particularly not to assume a perpetual growth rate greater than i . Even if one has difficulty in conceiving of a slow-down in the over-all growth rate g for any company (particularly *his* company), he could well admit this possibility with respect to some major segment of it. At least under a segment-based surplus theory, some slowly growing

or even declining segments can gain a competitive advantage from surplus previously built.

Magnitude of Surplus

This section seemed to bring forth less discussion than some of the rest, with the notable exception of Mr. Jackson's remarks. I am right with him when he argues that "no surplus can be sufficient to preserve the company against every possible catastrophe." In the event of a major atomic war, or some other catastrophic event seriously affecting the entire industry, the chances are that government assistance would be called upon and that the size of any particular company's surplus might turn out to be relatively unimportant. I get the impression that Mr. Jackson might be satisfied with a little less surplus than has been traditional among United States mutual companies. If this is his viewpoint, I am inclined to agree.

Mr. Jackson looks upon surplus primarily as a hedge against asset depreciation and hence asset-related; he attributes this view to the product-mix of the company with which he is associated. The author tends to bring the insurance risk into more or less equal prominence, at least partly because his company is heavily engaged in group insurance, issued on essentially a YRT basis, and therefore building up relatively little in the way of assets. The group insurance viewpoint is poorly represented among those who chose to discuss this paper; perhaps readers incorrectly inferred that the paper was primarily about surplus theory for individual insurance. As a matter of fact, the author's attempts to reconcile surplus objectives of the individual and group branches within his own company had much to do with the paper's having been written.

There is at least one suggestion, from Mr. Hill, that the optimum amount of surplus is not the most important thing; instead, the competitive aspects might be given priority in pricing, the growth objectives then being brought into line so that adequate surplus results. The interrelationships between growth, pricing, and surplus are obviously important in the practical situation, though for theoretical purposes the author likes the concept of setting surplus objectives first.

One or two of the others seem to object to considering surplus objectives first, then attempting to reach such objectives through the pricing structure, as is at least implied by the organization of the paper. In particular, Mr. Lebourveau, with his strong emphasis on adequate assets rather than adequate surplus, dislikes the idea of pricing being in any way dependent on growth. The author is not too happy with this concept either, but in some ways it is inherent in the arithmetic. There is, of

course, the alternative of a reducing surplus in times of unusual growth. This may be more satisfying to many people.

Sources of Surplus

Two of the discussers point out that there are alternates to viewing the interest on surplus as a source of surplus. Mr. Hill suggests interest on surplus as a source of agency expansion. Mr. Dorn suggests adding interest on surplus to other investment income and spreading the total in pricing calculations. Obviously, one can use any financial resource in any one of several different ways, but the over-all result is not necessarily affected if the goals remain unchanged. If Mr. Hill prefers to finance agency expansion out of interest on surplus, he reduces the contribution to surplus from its own interest but simultaneously increases the contribution to surplus from the insurance line no longer charged with agency expansion. To the author this is a distinction without a true difference, though he concedes that results to policyholders can be affected.

Mr. Cannon rightly points out that interest on surplus is heavily taxed. From this point of view, interest on surplus is not as good a source for surplus growth as one might otherwise assume.

Mr. Phillips goes on the line for a direct charge for surplus in the dividend calculation, as opposed to using somewhat conservative factors in the dividend formula. The author finds himself in agreement, essentially for the reasons Mr. Phillips gives.

Considerations of Equity

The author feels that he has not communicated very well with Mr. Dorn, who seems to feel that the paper's deliberate de-emphasis of legal considerations somehow undermines an important series of legal decisions. In particular, he indicates that the policy-based and company-based concepts of surplus allocation are somehow inconsistent with the concept of "dividend class" traditionally employed in individual insurance.

Perhaps we should here re-emphasize that the paper does not concern itself with surplus allocation in the Part 7 sense. The concepts of policy-, segment-, and company-based contributions to surplus all relate to any amounts held back *after* dividends have been declared. In the calculation of the dividend itself all of the concepts familiar to Mr. Dorn can be employed, and the resulting dividends will meet the test that dividends per thousand be equal for "all similar policies issued at the same time . . . under the same conditions, at the same age." The company-based and segment-based contributions to surplus obviously fit very well within this legal framework, and the policy-based can too if termination dividends are permissible.

Mr. Dorn appears to be especially disturbed with the idea that, under the policy- or segment-based contributions to surplus outlined in the paper, the long-range contribution to surplus from each policy in one case, each segment in the other, is nil. He clearly prefers that every policyholder contribute something to surplus, even if he gets part of his past surplus contributions back through termination dividends. Essentially, I am in agreement with Mr. Dorn, in that I believe some part of surplus contribution should be company-based and therefore nonreturnable. The completely returnable theory is tenable, however, whether Mr. Dorn and Mr. Trowbridge like it or not, and any company with termination dividends has gone at least part way toward it. As a practical matter, nearly every company the author knows about expects some company-based or nonreturnable contribution to surplus. Many have some element of returnable surplus as well.

A good illustration is found in Mr. Lebourveau's description of an ordinary dividend theory based on what he calls a "dividend fund." The dividend fund is essentially a combined reserve-and-surplus account which flows back on termination of the policy in the form of cash value (or death claim) plus termination dividend. Any surplus within the dividend fund is returnable and in this sense policy-based. Note, however, that there is outside the dividend fund another contribution to surplus in Mr. Lebourveau's system—the specific charges for the catastrophe reserve. This portion of surplus is clearly company-based, if both the individual policy and the cell to which it belongs leave an amount behind when the policy (or the cell) goes out of existence.

Note that if liabilities are defined in terms of the dividend fund, the surplus is reduced to the level of the catastrophe reserve. I might add that Mr. Lebourveau's formula looks very much like that of my own company. We view our dividend funds as reserves, for surplus analysis purposes. In effect they are gross-premium-valuation reserves, on realistic assumptions, with future annual dividends treated as one of the benefits. Note that there is no "surplus strain," in the sense of annual-statement surplus strain caused by net level premium reserves, if surplus is measured as the excess of assets over dividend funds. My point that dividends otherwise payable in early policy years may be low under a policy-based theory of surplus is not really related to the higher level of early expenses. Instead it comes about because the first-year contribution to surplus expected from a particular policy is large in respect to its parameters when the parameters are growing rapidly, particularly in the first year when the growth in the parameter is equal to the parameter itself. Suppose, for example, under a policy-based surplus theory, that each indi-

vidual policy were expected to furnish a surplus of 2 per cent of the face as a contingency reserve against the mortality risk. If the face amount is level, the entire 2 per cent is called for in the first policy year, with no further contribution thereafter. This is an extreme example of the author's concept that under a policy-based surplus theory "dividends otherwise payable are hit hard by the necessity of surplus accumulation."

Mr. Phillips gives us a good analysis of the development of annual-statement surplus under an individual level premium life policy. Because he is thinking in terms of net level premium reserves and early-year dividends, his policy-related surplus is likely to be negative at the early durations, and therefore another reason for surplus—to finance new business—enters Mr. Phillips' discussion. As stated earlier, the author's preference is to redefine surplus in other than statement terms, thereby eliminating the surplus strain in its usual form. At least this leads to a simpler analysis.

Mr. Link defines three distinct layers of surplus, with the totality of all three layers measured by the excess of assets over realistic reserves on a gross-premium basis. In applying his concept to the group annuity line of a mutual life company, he describes a typical group annuity dividend formula. The reserve, deducted from the experience fund in the calculation of dividends, contains part of Mr. Link's three-layered surplus. Whatever part it does contain is presumably returnable, because the reserves under a terminated case presumably go to zero eventually. Note another element of Mr. Link's formula which is not returnable—the special risk charge. Whether his "adjustment to compensate on a credibility basis for a portion of monthly gains and losses" results in another element of nonreturnable surplus is not immediately clear.

Finally, to sum up this long reply to a long discussion, the author is grateful to those who have contributed to this difficult subject. There are several important matters alluded to but not resolved, both in the paper and in the discussion. A few of these are (1) the relationships between growth objectives and surplus objectives, (2) Mr. Lebourveau's challenging statement that a mutual life insurance company has no owners and that its surplus is therefore essentially different from the net worth of the typical corporation with stockholders, and (3) Mr. Hazelcorn's challenge to equitable principles in the profit-making corporation, which the author assumed (and still believes) is maintained by the principle of a marketable ownership share.

These matters get right to the heart of the question of the reason mutual companies exist and whether they do in effect really differ from stock companies. There is clearly plenty of room for another author to tackle some of these "fascinating-but-elusive" questions.