

人事評価における行動情報の処理過程

——とくに評定誤差との関係について——

境 忠 宏

1. 問 題

人事評価は、処遇管理のみならず、適正配置のための人事情報源として、教育ニーズの発見や教育効果測定の評価基準として、あるいは能力開発やモチベーション管理のためのフィードバック情報として、人事管理の広範な領域において重要な役割を果たすものである。このような役割の遂行のためには、評価範囲や評価方法などにおいて、個々の管理目的に応じた評価制度（目的別評価）の確立が必要であるが、人事評価が人事管理の広い領域においてより有効に活用されるためには、個々の管理目的との斉合性ととも、人事評価そのものの信頼性と妥当性が十分に検証され、かつそれに係わるすべての人々に納得されるものでなければならない。

人事評価の信頼性に関して、もっとも大きな問題は様々な評定誤差（rating errors）の存在である。評定誤差は必ずしも意図的なものではなく、人の社会的情報の処理過程（cognitive process of social information）の基本的特質を反映するものであり、人が人を知覚したり評価する際に必然的に伴うものである。従来の評定誤差への対応方法としては、評価尺度の開発あるいは評定者訓練の実施などにより評定誤差の発生そのものを抑制しようとするものであったが、これらの効果についてはなお十分な検証はなされていない。一方、近年では、生じている評定誤差のレベルを把握するための人事評

価そのものの評価基準の開発や人の社会的情報の処理過程の理解とそれに応じた評価制度の設計による評定誤差の人事意思決定への介入の回避が重視されている。しかし、人事評価の評価基準を設定するためには、現在の評定誤差の定義自体がなお不明確であり、そのための測度も十分とはいえない。

本研究では、上記のような問題意識のもとに、まず評定誤差の種々のタイプを明確にするとともにその測度を提案する。ついで、これらの評定誤差の発生を規定する要因に関する従来の研究を展望するとともに、それらの要因が人の社会的情報の処理過程にいかなる影響をおよぼすのかを明らかにする。それによって、人の社会的情報の処理過程と評定誤差との関係についてのひとつの考え方を提示し、これを実験的に検討する。

2. 人事評価における評定誤差

人事評価に限らず、人物評価（person perception）全般におけるバイアスの存在を指摘したのは Thorndike(1920)と Kingsbury(1922)である。Thorndikeは、人は本来多面的でそれぞれの側面について異なる特徴をもつにもかかわらず、知覚者あるいは評定者がこれらの側面の相違を無視してしまい対象人物を全般的にすぐれている（あるいは劣っている）と判断してしまう傾向を指摘し、これをハロー（halo）効果と呼んでいる。また、Kingsburyは、ハロ

一効果以外に、他者をより寛大に評定する、あるいはより厳しく評定するという評定者の個人的特性や対象人物間の差別化そのものを回避しようとする傾向の存在を指摘している。これらは、現代では寛容性 (leniency) 効果および中心化傾向 (central tendency) あるいは尺度範囲の制限 (restriction of range) と呼ばれるものである。

Thorndike (1920) および Kingsbury (1922) 以来、これらの評定誤差に関する研究は数多いが、評定誤差そのものの概念的な定義はなおあいまいであり、かつ統一的な測度が設定されていないため研究者間での混乱がみられる。本研究では、まず、これらの定義を整理し、その測度について検討を加える。

ハロー効果 (Halo effect)

ハロー効果は、評定者が本来は相互に独立しておりかつ異なるはずの評価対象人物の種々の側面に対する弁別能力を欠いていることによるものであるが、その定義については研究者によって次のような混乱がみられる。

(1) 対象人物の全般的な印象あるいは対象人物への感情的態度が、対象人物の個々の側面についての知覚や評価を支配する傾向 (Borman, 1975 など)。

(2) 評定者の、対象人物の多様な側面に対する弁別能力の欠如 (DeCottis, 1977 など)。

(3) 対象人物をそれぞれの側面について齊合的に知覚・評価しようとする傾向 (Bernadin, 1977 など)。

これらの定義はいずれも、対象人物に対する知覚内容や評定結果の個々の側面間での強い相関関係の存在を示唆するものであるが、このような評定誤差の発生因については異なる考え方をとっている。まず、(2)は、ハロー効果は、人の情報処理能力の低さによるものとする点で、(1)および(3)の定義と大きく異なっている。また、(1)と(3)は、ともに、人の情報処理

様式のもつ特質をハロー効果の発生因としているが、(1)は、人の情報処理におよぼす感情的側面の影響力をその基本的源泉とする点で(3)と異なる。(3)は、人物知覚に限らず、人の知覚過程そのもののもつ基本的特質、つまり激しく変化する外部環境をより安定的で予測可能なかたちで知覚し、認知しようとする認知的体制化傾向をハロー効果の発生因とするものである。このような、ハロー効果の定義そのものについての概念的混乱に対して、Nisbett と Willson (1977) は、発生のメカニズムが異なるふたつのタイプのハロー効果の存在を指摘している。ひとつは、全般的印象が、個々の側面についての知覚や評価そのものを変容し歪曲してしまうという強いハロー効果であり、ひとつは、全般的印象があいまいな知覚の手掛りの解釈や欠如している情報の推測に一定の方向性を与えるという弱いハロー効果である。また、Nisbett らは、強いハロー効果の存在は実験方法によってのみ例証可能であり、日常的場面ではむしろ弱いハロー効果の方が一般的であるとしている。われわれも、人事評価におけるハロー効果は、Nisbett らのいう弱いハロー効果によるものと考える。これは、基本的に人の情報処理様式の基本的特質を反映するものであり、評価方法の開発や評定者訓練だけでは対応することのできないものである。むしろ、職務分析の徹底による職務要件の明確化や評価対象範囲の制限、あるいは評価目的の特定と評価項目の具体化など評価システム全体の見直しが必要であり、そのためには、人の行動情報の処理様式そのものの十分な理解を必要とする。ハロー効果発生背後にあると思われる人の認知過程の特質については5章において詳細に検討する。

ハロー効果の測度としてしばしば用いられるのは、個々の側面についての評定結果間の相関係数であり、相関係数が大きいほど個々の側面についての評価の分化度は低く、より強いハロー効果が存在するとされる。また、これをより発展させたものとして、因子分析を用いた研究

も多い。つまり、抽出される因子数が少なく、かつそれによって説明される変動分の比率が大きいほど、評価結果の分化度はより低く、ひとつの因子によってほとんどの変動が説明されるような場合にハロー効果は最大とされる。

これらの測度は、ハロー効果の概念的定義と密接に係わるものであり、その意味するところは直観的にわかりやすいが、測度そのものは評価結果にみられる構造的特性の指標にすぎず、必ずしもハロー効果による評価のバイアスのレベルそのものを表わすものではない。たとえば、評価対象とされた側面間の共変関係そのものが強ければ、たとえハロー効果による評定誤差が存在しなくともこれらの測度は大きくなる(因子分析を用いる場合には、抽出される因子数は少なくなる)。したがって、ハロー効果による評定誤差を検出するためには、評価結果全体の信頼性が検討できるとともに、そのなかでのハロー効果によるバイアスのレベルを把握することのできる測度の開発が必要である。このような測度については本章の最後に提案する。

寛容性効果 (Leniency effect)

寛容性効果は、特定の評定者が、対象人物全体をより高い方向に(あるいはより低い方向に—severity 効果)評価してしまうという傾向である。したがって、ハロー効果が、特定の対象人物の異なる側面についての弁別力を低下させるのに対して、寛容性効果は、特定の評定者の対象人物間での弁別力を低下させるものである。このような評定誤差については、次のようないくつかの定義が存在する。

(1) 実際のレベルよりも、他者をより高く(あるいはより低く)評価しようとする一般的な傾向 (Saal と Landy, 1977 など)。

(2) 人物評価における他者への要求水準のレベルという評定者の個人的特性の反映 (DeCotiis, 1977 など)。

(3) すべての対象人物についての評価結果の

平均値の尺度上の中位点 (the midpoint of the rating scale) からの逸脱 (Benardian, 1977 など)。

これらの定義のなかで、(1) は寛容性効果を、人々の一般的な評価傾向の表われとするものであるが、このような評価傾向が発生する心理機制が明確にされていないとともに、leniency と severity の発生方向を規定する条件も確定されていない点で不十分である。また、本来測定不能な真の評価レベルの存在を仮定しているという点で大きな問題をはらんでいる。また、(3) は、寛容性効果による評定誤差のあらわれ方そのもので寛容性効果を定義するという誤りを犯しているが、同時に、尺度上の中位点を実際の評価分布における中性点と同一とみなすという点でやはり大きな問題をはらんでいる。このなかでは、(3) が、寛容性効果をその発生因によって定義しているという点でもっとも妥当なものと思われるが、他者評価における要求水準のレベルそのものの規定因が明確にされていないという点ではなお不十分さが残る。

われわれは、寛容性効果は、人が他者評価において自らの自己評価を評価基準として設定することによる一種の対比効果および同一化効果の表われと考える。したがって、寛容性効果の出現を規定するひとつの要因は、評定者自身の自己評価のレベルという評定者の個人的特性にある。この点で、われわれは、(2) の定義に基本的に同意する。しかし、対比効果が生じるか(したがって、severity 効果が生じるか)同一化効果が生じるか(したがって、leniency 効果が生じるか)は、対象人物と評定者との心理的距離に依存し、対象人物との心理的距離が小さな場合には同一化効果が生じやすく、心理的距離が大きな場合には対比効果が生じやすいものとする。また、対象人物との心理的距離の規定因としてもっとも大きな要因は、両者の集団成員性の異同であり、同じ集団に所属している人物との心理的距離ほど小さく、寛容性効果がより生じやすいものと思われる。このため、管

理者による職場の部下に対する人事評価においては、severity 効果よりも leniency 効果の方がより出現しやすいことになる。この点で、われわれは、(2) の定義をこえ、寛容性効果は、自己評価という個人的特性とともに (1) の定義が示唆する人々の一般的な他者評価傾向との交互作用効果によるものであるという視点をとることになる。なお、このような寛容性効果が、人の社会的情報処理様式の基本的特質といかにかかわっているかについては、5章で詳しく検討する。

寛容性効果の測度としてよく用いられるのは、個々の側面についての評定値の平均の尺度上の中位点からの逸脱であり、平均が中位点以上であれば leniency 効果が、中位点以下であれば severity 効果が生じているとされる。しかし、この測度は、(3) の概念的定義における問題点として指摘したように、尺度上の中位点を評価分布上の中性点と同一とみなすという点で妥当なものとはいえない。この点では、評定値分布の左右へのかたよりを示す歪度 (skewness) を測度とする方がより妥当であり、統計的に有意な正の歪度は leniency を、負の歪度は severity を反映していることになる。なお、歪度は次式によって定義される。

$$g_1 = \frac{u_3}{u_2 \sqrt{u_2}}$$

ここで、 g_1 は歪度を、 u_2 は2次のモーメントを、 u_3 は3次のモーメントを表わす。また、 u_2 、 u_3 は、データ数を N とし、それぞれの評定値の平均値との差を d_i とすると、

$$u_2 = \sum_{i=1}^N d_i^2 / N$$

$$u_3 = \sum_{i=1}^N d_i^3 / N$$

となる。

しかし、この測度も、ハロー効果の一般的な測度と同様に、評定値全体のもつ信頼性との関係で寛容性効果による評定誤差のレベルが検出されていないという点では、評価そのものの評

価基準としてはなお不十分である。

中心化傾向 (Central tendency)

中心化傾向は、評定者が他者に対する極端な評価を回避し、尺度上の中位点付近しか用いないという傾向である。しかし、尺度上の中位点は必ずしも評価的な中性点を表わさず、しかも、leniency や severity により評価分布の中央が尺度上の中位点の左右いずれかにかたよる以上、これはむしろ尺度範囲の制限 (restriction of range) と呼ぶ方が妥当である。

中心化傾向については、評定値分布の特定範囲への集中という点では一致しており、その分布の中央が尺度上の中位点となるのかそこからずれるのかという点でのみ概念上の混乱がみられる。この点では、尺度範囲の制限がより一般的な評定誤差のタイプであり、中心化傾向はその1ケースにすぎないといえる。そのとき、尺度範囲の制限は、評定者が尺度上の極端な部分 (極性点付近) の利用を回避するわけではなく、対象人物間の評価差の拡大を回避する傾向として定義されることになる。なぜなら、leniency 効果が生じているような場合には、正の方向の極性点は利用されることになる。また、このような尺度範囲の制限は、人物評価における不確実性の増大を回避しようとする傾向のあらわれであり、やはり環境を安定的でより予測可能なものとして知覚しようとする人の情報処理様式の特徴を反映するものと思われる。

中心化傾向あるいは尺度範囲の制限の測度としてよく用いられるのは、評定平均の尺度上の中位点との近接性 (中心化傾向) あるいは特定の側面についての対象人物個々に対する評定値全体の標準偏差 (尺度範囲の制限) であり、尺度上の中位点との近接性が高いほど中心化傾向は強く、標準偏差が小さいほど尺度範囲の制限は強いとされる。また、寛容性効果と同様に、評定値分布の形態的特性に注目し、分布の尖度 (kurtosis) をその測度とする研究もある。つま

り、分布の尖度が大きいほど、尺度範囲の制限は強いとするものである。なお、尖度は次式によって定義される。

$$g_3 = \frac{u_4}{(u_2)^2}$$

ここで、 g_3 は尖度を、 u_2 は2次のモーメントを、 u_4 は4次のモーメントを表わす。

しかし、この測度も、ハロー効果や寛容性効果と同様、評定結果全体の信頼性との関係については何も明らかにしていない。

評定誤差については、これら3つ以外にも、人々のもつ人間性についての常識的な仮説(implicit personality theory)にもとづく論理的錯誤(logical error, Newcomb, 1931)や評定項目の空間的な位置関係にもとづく近接効果(proximity error, Stockford と Bissell, 1949)あるいは行動の観察順序や情報との接触順序にもとづく順序効果(first and last impression error, Latham, Wexley と Pursell, 1975)などがあるが、本研究ではこれらの評定誤差についてはとりあげない。

3. 人事評価の評価基準

前章で指摘したように、評定誤差の従来の測度について共通の問題点は、それらと評定結果全体の信頼性との関係が不明確であり、人事評価の評価基準としての要件を十分に満たしていないというところにあった。

人事評価の信頼性(あるいは妥当性)は本来、評価結果と何らかの客観的指標との相関関係によって検証されるべきものである。しかし、人事評価が利用される理由のひとつは、個々人の能力や行動の客観的な指標の欠落あるいは利用の困難さにあり、したがって、このような検証方法そのものがきわめて困難なことになる。このとき、人事評価の信頼性の評価基準として用いるのは、異なる評定者間での評定結果の一致(interrater reliability or agreement)であ

る。しかし、評定者間での一致それ自体は必ずしも評定結果の信頼性の指標にはなりえない。なぜなら、評定者それぞれが“正しい”評定はしていなくても、全員が同じ誤まりを犯していれば評定者間での一致そのものは高まることになる。したがって、複数の評定者による評定結果全体のなかに評定誤差が生じておらずあるいは生じていてもそのレベルは低く、かつ、評定者間での一致度が高いときのみ、評定結果は十分に信頼しうる妥当なものと結論することができる。

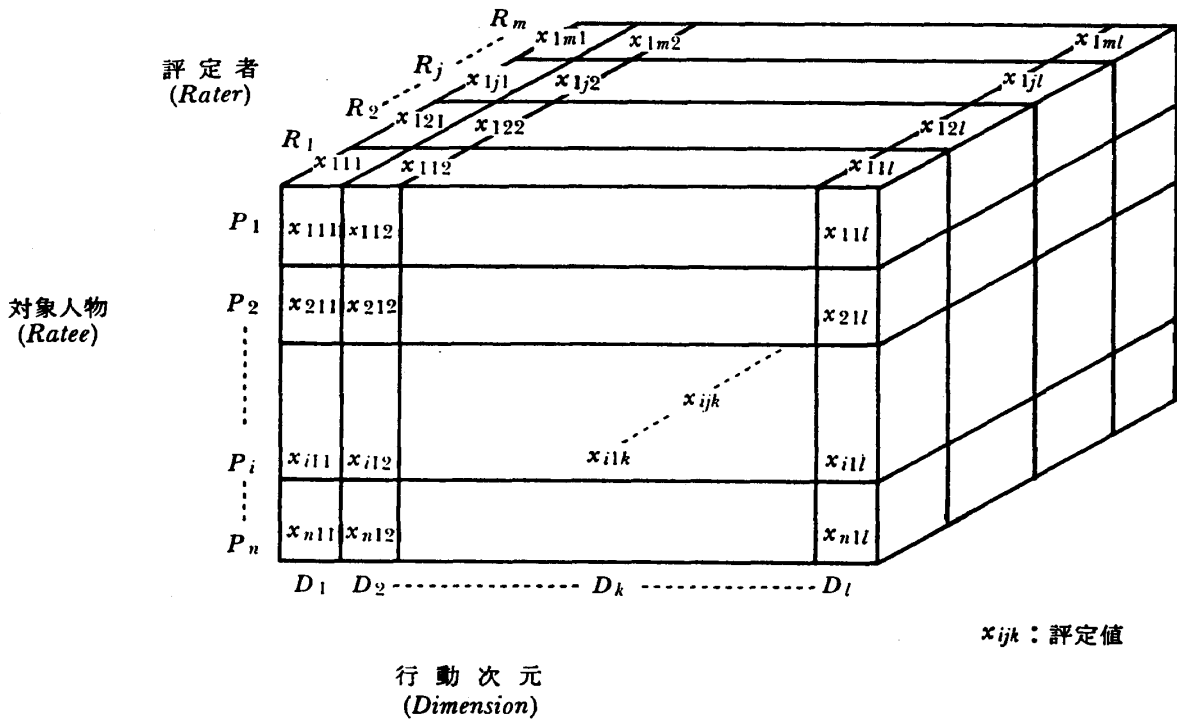
ここから、人事評価の評価基準としては、評定者間での評価の一致度とともに、それとくらべた評定誤差のレベルという複合基準が必要になることになる。このような評価基準が適用されるには、人事評価は図1に示すような構造をもたねばならない。このデータ構造のもとでのみ、評定者間での一致度と種々の評定誤差のレベルとが統一的な測度によって比較可能となる。

このような構造のもとでは、評定者間での一致度および3つの評定誤差は分散分析における種々の効果指標を用いることによって、定義することができる。まず、図1のデータ構造から次のような平均平方(Mean Square)を算出する。

(1) 特定の行動次元(dk)についての評定者間変動の平均平方。これを、 $MSR(dk)$ と表記する。なお、 $MSR(dk)$ の算出式は次のようなものとなる。

$$MSR(dk) = \frac{\left(\sum_{i=1}^n x_{i1k}/n - \sum_{j=1}^m \sum_{i=1}^n x_{ijk}/m \cdot n \right)^2}{m-1} + \frac{\left(\sum_{i=1}^n x_{i2k}/n - \sum_{j=1}^m \sum_{i=1}^n x_{ijk}/m \cdot n \right)^2}{m-1} + \dots + \frac{\left(\sum_{i=1}^n x_{ijk}/n - \sum_{j=1}^m \sum_{i=1}^n x_{ijk}/m \cdot n \right)^2}{m-1}$$

図1 人事評価の評価基準に必要な構造 (データ構造)



x_{ijk} : 評定値

(2) すべての行動次元についての評定者間変動の平均平方。これを、 MSR と表記する。 MSR の算出式は次のようになる。

MSR

$$= \frac{\left(\sum_{i=1}^n \sum_{k=1}^l x_{i1k} / l \cdot n - \sum_{j=1}^m \sum_{i=1}^n \sum_{k=1}^l x_{ijk} / l \cdot m \cdot n \right)^2}{m-1}$$

$$+ \frac{\left(\sum_{i=1}^n \sum_{k=1}^l x_{i2k} / l \cdot n - \sum_{j=1}^m \sum_{i=1}^n \sum_{k=1}^l x_{ijk} / l \cdot m \cdot n \right)^2}{m-1}$$

$$+ \dots$$

$$+ \frac{\left(\sum_{i=1}^n \sum_{k=1}^l x_{imk} / l \cdot n - \sum_{j=1}^m \sum_{i=1}^n \sum_{k=1}^l x_{ijk} / l \cdot m \cdot n \right)^2}{m-1}$$

なお、以下の平均平方についての算出式は省略する (同様な方法によって求めることができる)。

(3) 特定の行動次元についての対象人物間変

動の平均平方。これを、 $MSP(dk)$ と表記する。

(4) すべての行動次元についての対象人物間の平均平方。これを、 MSP と表記する。

(5) 特定の行動次元についての評定者と対象人物の交互作用変動の平均平方。これを、 $MSRP(dk)$ と表記する。

(6) すべての行動次元についての評定者と対象人物の交互作用変動の平均平方。これを、 $MSRP$ と表記する。

(7) 評定者、対象人物、行動次元の3変数間の交互作用の平均平方。これを、 $MSRPD$ と表記する。

これらの平均平方に基づいて、評価の信頼性の評価基準として次のような測度を設定することができる。

(1) 評価の一致度 (Agreement)

評定者間の評価結果の一致度の指標としては、級内相関 (intraclass correlation) を用いることができる。

したがって、

$$\text{Agreement} = \frac{MSR - MSRP}{MSR + (m-n)MSRP}$$

となる。

(2) ハロー効果による評定誤差 (Halo)

図1のような構造のもとでは、ハロー効果による評定誤差は、評定者・対象人物・行動次元の3要因のもとでの変動の大部分が、評定者と対象人物との交互作用によって説明されてしまい、行動次元の相違による変動が最小化するというかたちであらわされることになる。したがって、ハロー効果による評定誤差のレベルは、評定者・対象人物・行動次元の3要因間の交互作用変動にくらべた評定者と対象人物の2要因間の交互作用変動の比率としてとらえることができる。

ここから、

$$\text{Halo} = \frac{MSRP}{MSRPD}$$

となる。

(3) 寛容性効果による評定誤差 (Leniency)

寛容性効果は、特定の行動次元における特定の評定者の評価傾向の正方向あるいは負方向へのかたよりであり、図1の構造のもとでは評定者要因の主効果としてとらえることができる。

したがって、

$$\text{Leniency} = \frac{MSR(dk)}{MSRP(dk)}$$

となる。なお、この指標は、それぞれの行動次元ごとに算出されねばならない。

(4) 中心化傾向あるいは尺度範囲の制限による評定誤差 (Restriction of Range)

中心化傾向あるいは尺度範囲の制限による評定誤差は、対象人物に対する評価が一定の範囲に集中することによる評価の対象人物間での分化度の低下としてとらえることができる。したがって、図1の構造のもとでは、中心化傾向あるいは尺度範囲の制限による評定誤差は、それぞれの行動次元における対象人物要因の主効果の欠落としてとらえることができる。

ここから、

$$\text{Restriction of Range} = \frac{MSP(dk)}{MSRP(dk)}$$

となる。なお、この指標も、それぞれの行動次元ごとに算出されねばならない。また、この指標のみは、他の指標と異なり、数値が小さいほど（したがって対象人物要因の主効果が小さいほど）評定誤差は大きなことを表わしている。

これらが、人事評価の統一的な評価基準として用いられるべき測度であり、人事評価に次のような要件の充足を求めるものでもある。

- (1) 評定者間での有意な一致が存在すること。
- (2) それぞれの行動次元での評価が有意に分化していること。
- (3) それぞれの評定者に有意な評価のかたよりがみられないこと。
- (4) それぞれの対象人物についての評価が有意に分化していること。

4. 評定誤差の規定因

前章において3つの評定誤差を定義し、評価の信頼性に関する評価基準を設定したが、ここでは人事評価の過程に係わる要因のなかでとくにどのようなものが評定誤差の出現を強め、評価の信頼性をそこねるのかという視点から従来の研究を展望する。

評定誤差の出現に係わる要因として従来の研究において検討されているものは、評価に用いられる尺度の形式・評定者訓練の有無・評定者の特徴・評価対象人物の特徴や行動などである。本研究では、これらの要因についての従来の研究結果を簡単に紹介するとともに、評定者の社会的情報の処理過程という視点からこれらの結果を統合的に理解するための基本的枠組を提示する。

(1) 尺度形式

評価に用いられる尺度の形式と、評価の信頼性および評定誤差との関係を体系的に検討した

のは, Barret, Taylor, Parker と Martens(1958)である。Barret らは, 次の4つのタイプの図式評定尺度(graphic scale)を用い, それぞれの尺度のもとでの評価の信頼性と評定誤差を検討している。

Format I. 15段階からなる評定尺度(10インチの線分を15の区分に区切ったもの)に評価すべき特性名のみを付したのもの。

Format II. 特性名を提示するだけでなく, 特性の内容の詳しい説明を付加したもの。

Format III. 特性名を提示するとともに, 線分上のそれぞれの段階に対応する行動基準(behavioral anchors)を明示したもの。

Format IV. 特性名は提示せず, 線分上のそれぞれの段階に対応する行動基準のみを明示したもの。

Barret らの結果では, これらのなかでは, Format III を用いた場合に, 評価の信頼性が高まり, 寛容性効果やハロー効果による評定誤差は弱まることが見い出されている。この研究以来, 評定尺度における具体的な行動事例の提示や行動基準の明示が重視され, 行動基準評定尺度(Behaviorally Anchored Rating Scale, BARS)が開発されている。この尺度は, 線分上の評価ポイントが, 特性形容詞や数値ではなく, 具体的な行動事例によって示され, かつこれらの行動事例は等時間隔法によってその評価値を決定され, この評価値に基づいて線分上に配列されている。従来の図式評定尺度(Barret らの Format I あるいは Format II)と行動基準評定尺度との比較では, 評価の信頼性の高さおよび評定誤差のあらわれにくさという面では行動基準評定尺度の優位性が見い出されている(たとえば, Borman と Dunnette, 1975, Keaveny と McGann, 1975 など)。とくに, Keaveny らは, 行動基準評定尺度を用いた場合には, ハロー効果があらわれにくいことを示している。しかし, Borman らは, 行動基準評定尺度は信頼性や評定誤差という面で図式評定尺度よりもすぐれているが, それでも尺度形式

の相違は評定結果全体の変動の約5%しか説明できないことを示している。

行動基準評定尺度は, 評価さるべき行動次元を特定し, その手掛りとすべき行動事例を明示することでハロー効果の発生を抑制しようとするものであるが, Keaveny らが示すように, この尺度はハロー効果の抑制には有効であっても寛容性効果からはのがれていない。寛容性効果を避けるためには, 評定における社会的価値判断(social desirability)そのものが統制されなければならない。このような考え方から開発されたものに, 強制選択尺度(forced choice rating scale)がある。この尺度では, やはり事前に評価値の決定された行動事例が用いられるが, それが評価値に沿って配列されるのではなく, 評価値の等しい行動事例が4つずつ提示され, 評定者はそのなかから評価対象人物にもっとも適合すると考える行動事例をひとつだけ選択することになる。この尺度と図式評定尺度とを比較し, Taylor, Schneider と Clay (1954)は, 強制選択尺度を用いた場合には, 寛容性効果が著るしく弱まることを, また, Cotton と Stoltz (1960)は, 尺度範囲の制限が生じにくくなることを見い出している。これらの研究から, 強制選択尺度は, 寛容性効果や尺度範囲の制限を弱め, 評価結果の対象人物間での分化度を高めるには有効であるといえるが, 対象人物内の行動次元間での評価の分化度に対する効果については明らかではない。

このように, 尺度形式と評定誤差との関係については多数の研究があるが, これらの研究に共通する問題点は, 尺度がそれぞれの評定者の判断を表明する手段としてしかとらえられていないというところにある。しかし, 人事評価は基本的に記憶に基づいた判断であり, 過去に観察された行動情報の記憶過程における認知的体制化こそがハロー効果や寛容性効果を生み出す基本的源泉となっているはずである。このような認知的体制化現象の詳細については次章で述べることにするが, この視点から評定尺度に要

求されることは、評定尺度と人々が行動情報の記銘・把持・想起に用いる認知構造 (cognitive structure) との適合関係である。つまり、尺度形式そのものよりも、それが行動情報の記銘時における人の認知構造の分化度にどれだけ貢献しているか (評定尺度が設定しているような行動次元が観察者の認知構造のなかにも確立されているかどうか)、あるいは評定尺度が過去において観察された行動のどれだけ適切な想起手掛りとなりえているか、また、評定尺度の用意しているカテゴリーが想起された行動情報を統合するのにどれだけ有効な道具となりえているかという視点からの検討がより重要である。

この場合、具体的な行動事例や行動基準の明示は確かに評価の要求されている行動次元あるいは想起すべき行動情報を特定するうえで図式評定法にくらべるとすぐれているといえる。しかし、たとえ図式評定法であっても、そこで設定されている次元が評定者の認知構造に適合するものであれば、行動基準評定尺度と同様な結果を生み出しうるはずである。Friedman と Cornelius (1976) は、行動基準評定尺度の開発に参加したグループ、図式評定尺度の開発に参加したグループ、いずれの尺度の開発にも参加しなかったグループの評定結果を比較し、第3のグループでは、第1のグループや第2のグループとくらべいずれの評定尺度を用いた場合にもハロー効果による大きな評定誤差が生じているが、第1のグループと第2のグループ間では差のないことを見い出している。これは、評定誤差が尺度形式そのものに依存するわけではなく、尺度形式と評定者の認知構造との適合化の欠如によるものであることを示している。

また、行動基準評定尺度では、用いられるカテゴリー数が増加し、尺度そのものが複雑化するが、人の情報処理容量の限界から考えれば、これは評価の信頼性を低下させる大きな原因ともなる。たとえば、Finn (1972) は、評価に用いられるカテゴリー数と評価の信頼性との関係を分析し、カテゴリー数が2つ以下あるいは8

つ以上になると評価の信頼性は著るしく低下することを示している。

したがって、評定尺度の設計において重要なことは、形式そのものではなく、評定尺度において設定される次元の内容や次元の数と評定者の用いる認知構造との適合化にあるといえる。これは、評定尺度の認知構造への適合化のみでなく、評定者訓練による評定者の認知構造の評定尺度への適合化によっても可能である。なお、前者の場合には、評定者の認知構造そのものの把握が必要となるが、そのための分析手法および認知構造の構造的特性と人の情報処理様式との関係については、境 (1985) を参照されたい。

(2) 評定者訓練

評定者訓練が評定誤差の低減におよぼす効果については、研究結果は必ずしも一致していない。評定者訓練の評定誤差の低減効果を報告している研究も多いが、Vance, Kuhnert と Farr (1978)などは、訓練を受けた評定者と訓練を受けていない評定者間で、評価結果の信頼性や評定誤差に差はないことを見い出している。また、Bernadian (1978) は、評定者訓練により評定誤差の低減効果はみられても、それは訓練直後にのみみられる短期的なものであり、この効果も時間経過にともない解消されることを報告しており、Borman (1975) は、評定者訓練はハロー効果のある程度の低減はもたすが、評価の信頼性そのものには影響しないことを見い出している。

このように評定者訓練の効果については、一致した研究結果が見い出されていないが、これは評定者訓練そのものが十分に構造化されていないためと思われる。つまり、評定誤差の低減に必要なことは、評定尺度において評価の要求されている行動次元が評定者の認知構造のなかにも明確に設定されていることとこのような評定誤差を生み出す自らの情報処理過程についての十分な自覚が存在することであるが、従来の評

定者訓練ではこのような視点が欠落しているように思われる。むしろ、評定者訓練以外の要因をとりあげている研究において、認定誤差の低減に必要な評定者の要件が明らかにされている。

たとえば、尺度形式の影響でもとりあげたように、Friedman と Cornelius (1976) は、評定尺度の開発への参加がハロー効果の著るしい低減をもたらすことを示しており、これは評定尺度の要求する評価次元と評定者の認知構造との適合化が認定誤差の大きな低減効果をもつことを示唆するものである。また、Gordon (1970) は、評定者訓練経験よりも、特定の評定尺度の使用経験の方が評価結果の信頼性をより強く規定していることを、Klieger と Mosel (1953) は、評価対象人物に要求されている職務要件に対する理解度の高さが評価の信頼性を強く規定していることを見い出している。これらはいずれも、評定者の認知構造の評定尺度への適合化や認知構造そのものの分化度の向上が評価の信頼性に大きく影響していることを示すものである。あるいは、Bernadian (1978) は、認定誤差に関する知識や理解度と実際に生じる認定誤差との相関関係を検討し、認定誤差に対する理解が深まるほど生じる認定誤差も弱まることを見い出している。これは、認定誤差の発生過程に対する理解と自覚の評定誤差そのものの低減効果を示すものであろう。

したがって、評定者訓練は、単に一般的な評定尺度に関する知識の提供や行動観察技法の学習のみではなく、特定の評価状況における特定の評定尺度と評定担当者の認知構造の適合化および人事評価における自らの行動情報の処理過程に対する理解と自覚を引き出しうるにより構造化されねばならない。

(3) 評定者の特徴

評価の信頼性と認定誤差に影響をおよぼす評定者の特徴としては、評定者の認知的複雑性 (cognitive complexity)、評定者自身の自己評

価および対象人物との心理的距離 (psychological distance)、評定者のリーダーシップ・スタイル、職務および対象人物についての評定者の知識などが見い出されている。

このなかでも、とくに影響力が強く、かつその方向が明確なのは、評定者の認知的複雑性である。認知的複雑性は、人が知覚対象を分類したり、情報を記銘・処理する際に用いる認知構造の構成次元数を表わすものであるが、Schneier (1977) は、認知的により複雑な評定者ほど、ハロー効果や寛容性効果は弱く尺度範囲の制限も生じにくいことを示している。また、認知的複雑性の高い評定者ほど、図式評定尺度よりも行動基準評定尺度の方を用いることを好むことも見い出されている。ここから、人事評価における認定誤差の出現を強く規定しているのは、評定者の情報処理能力や情報処理様式であり、尺度形式はこのような評定者の情報処理様式に適合する場合にのみ認定誤差の低減効果をもつといえる。

Mandell (1956) は、自己評価の高い評定者ほどより寛容性効果は強まることを示しているが、Rothaus, Morton と Hanson (1965) は、対象人物との心理的距離が大きな場合にはむしろ評価は厳しくなる (severity 効果が生じる) ことを報告している。これは、寛容性効果 (あるいは severity 効果) が、基本的には自己評価との同一化 (あるいは対比) によるものであり、そのいずれが生じるかを規定するのは対象人物との心理的距離であることを示すものである。また、Klores (1966) は、評定者のリーダーシップ・スタイルと認定誤差との関係を検討し、人間関係指向的な (high in consideration) リーダーほど寛容性効果が強く、課業指向的な (high in initiation of structure) リーダーほど、尺度範囲の制限は弱く、かつ課業遂行に関連した行動次元上での対象人物評価の分化度がより大きいことを示している。これは、評定者 (管理者) の行動様式自体が、評価に用いる情報 (行動的手掛り) の観察範囲を規定するとと

もに注意を方向づけるためと思われる。

Amir, Kovarsky と Sharan (1970) は、評価対象人物の担当する職務の内容や要件についての知識があるほど評価の信頼性は高まることを見い出しているが、対象人物自体についての知識は評価の信頼性には影響しないことを示している。また、Landy と Guion (1970) は、評価の信頼性を規定するのは、評価の対象となる行動の観察経験であり、対象人物との日常的な接触経験と評価の信頼性との関係は弱いことを示している（前者の条件では、信頼性係数は0.62であるが、後者の条件では0.24にすぎない）。ここから、評価の信頼性を規定するのは、対象人物自体についての知識ではなく、観察すべき行動次元についての理解やその次元に係わる行動的手掛りの保有度であるといえよう。

したがって、評価の信頼性や評定誤差に影響するのは、評定者の情報処理能力や情報処理様式であり、認知構造の分化度が高く、観察すべき行動次元についての理解度が高く、かつその次元に関する行動的手掛りが多いほど、評定誤差は弱く、評価の信頼性は高まることになる。

(4) 対象人物の行動

Hamner, Kim, Baird と Bigoness (1974) は、人事評価の変動の最大の説明変数はやはり対象人物の実際の職務行動 (job performance) であるが、それでも職務行動は評価全体の変動の30%しか説明できないことを示している。また、Scott と Hamner (1975) は、職務成果の平均的な水準は一定でもその時系列的な変動が評価を強く規定していることを示している。たとえば、平均的な成果水準 (performance level) は同一でも、その時系列的な変動が大きい場合には、能力評価は高まるものの意欲評価は著るしく低下する。また、最初に高い成果をあげて、それが徐々に低下するような場合には、平均的なレベルは同一でも、評価は著るしく低下する。これらの結果は、人事評価は確かにある程度、対象人物の職務行動を反映するが、それ

以上に評定者の情報処理様式の特徴を反映していることを示すものである。

5. 人事評価における行動情報の処理過程

人事評価は、評定者による対象人物の行動観察から始まり、観察された行動情報の認知構造への組みこみ、記憶過程における行動情報の認知的体制化、評価時点における行動情報の記憶からの検出、検出された行動情報の統合という一連の情報処理過程から構成されるものである。このような人の情報処理過程の基本的特質は、環境の認知的単純化 (cognitive simplification of the environment) にある。つまり、人は、複雑な側面を有し時々刻々と変化する環境を、より分節化され構造化され、比較的安定したものとして知覚することで、自らの環境への適応性と支配力を高めようとする傾向を有している。また、これは、人間の情報処理能力の限界によるものであり、人は自らの感覚器官に到達する無数の情報をいちいち処理することによってではなく、あらかじめ設定された知覚的あるいは認知的枠組みのもとで情報を選択し、認知的枠組みに統合することで環境を知覚し理解しようとする。このような情報処理様式は、制約された情報処理能力のもので人間の環境適応能力を著るしく高めることになるが、同時に現実と知覚との間に種々のギャップを生み出すことになる。評定誤差は、このような現実の行動とそれについての知覚とのシステムティックなギャップあるいはバイアスによるものであり、本来の意味での誤差 (error) ではなく、人間の情報処理様式の基本的特質を反映するものである。

人間の社会的情報の処理様式に基づいた人事評価過程については、Feldman (1971) などにより詳細なモデル展開がはかられているが、ここでは、観察された行動 (observed behaviors) の記銘と想起の方向を強く規定するカテゴリー化について検討を加え、カテゴリー化 (cate-

gorization) という人の情報処理様式の特徴が評定誤差といかにかかわっているかを実験的に検討することにする。

Cantor と Mischell (1979) によれば、人は観察された行動そのものを記憶としてとどめるのではなく、観察された行動(あるいは対象人物の外見的特徴)を手掛りとして自らのもつ人物カテゴリーのいずれかへの対象人物の所属性(category membership)を判断する。このとき、所属性の判断基準となるのは、それぞれのカテゴリーの典型的事例(category prototype)であり、対象人物とプロトタイプとの類似性が所属性判断の基本的手掛りとなる。対象人物のカテゴリー化が成立すると、人はカテゴリー・プロトタイプと一致する行動情報へ注意を向けやすく、不一致な行動情報は無視されるようになる。また、不一致な行動情報が反復して出現する、あるいは出現頻度は少なくともそれがより顕在的な(salient)な場合には、注意の方向が意識的に統制され不一致な行動情報も短期的には記録されるようになるが、それがカテゴリー化の修正を導出する前に、行動原因の帰属判断(causal attribution)が行なわれる。

カテゴリー・プロトタイプは、行動事例の集合ではなく、より内面的な人格特性の集合であり、観察された行動の手掛りから対象人物の分類されるカテゴリーが判断されるとそのカテゴリー・プロトタイプのもつ人格特性集合が対象人物にも帰属されることになる。しかし、ひとたびカテゴリー化が成立したあとでは、観察された行動はその原因が対象人物の内面的特徴にあるのか(内的帰属と呼ぶ)対象人物の置かれている環境にあるのか(外的帰属と呼ぶ)が判定され、内的帰属判断が生じた場合にのみ行動情報は対象人物の内面的特性を示すものとして記録されることになる。このとき、既存のカテゴリー化と不一致な行動について内的帰属判断が生じる場合にのみ、カテゴリー化の修正あるいはカテゴリー構造そのものの再構成が行なわれることになるが、事前のカテゴリー化そのも

のも帰属判断の大きな規定因であり、カテゴリー・プロトタイプと一致する情報は内的原因に帰属判断されやすく、カテゴリー・プロトタイプと不一致な情報は外的原因に帰属判断されやすい。したがって、ひとたびカテゴリー化が成立するなら、カテゴリー・プロトタイプと一致する行動の手掛りのみが注目されやすく、かつ記録されることになり、たとえそれと不一致な行動の手掛りが存在したとしてもそれは知覚されにくく、また知覚されたとしても外的原因に帰属され、再カテゴリー化やカテゴリー構造の再構成(この場合には、認知構造を構成する次元数が増加するとともに、新たな階層構造が出現することになる)は生じにくいことになる。

また、カテゴリー化は、評価時点における行動情報の想起にも大きな影響をおよぼす。観察された行動の手掛りは短期間は記憶に把持されるが、長期的には、対象人物の分類されたカテゴリーとこのカテゴリー・プロトタイプのもつ特性集合のなかで対象人物に帰属される特性の下位集合のみが記憶に把持されるようになる。したがって、評価時点においては、まず、対象人物に帰属されたカテゴリーとそのプロトタイプが想起され、そのなかで対象人物に付与された特性集合が想起される。これを手掛りとして、これらの特性集合と合致した行動事例が再生されることになる。ただし、Wyer と Srull (1980) は、行動情報が直接的に想起されないのは長期記憶(long-term memory)が介在する場合であることを、また、Cohen (1981) は、一致情報をより想起しやすくなるのは情報過重が存在する場合であり、情報が過重でなければ、事前のカテゴリー化は一致情報とともに不一致情報をもより想起させやすくなることを示している。これは、事前カテゴリー化の存在は、情報の評価基準を与えることになり、一致情報とともに不一致情報の顕在性(salience)をもより高めるためである。しかし、情報量が増し、行動情報の直接的な把持が困難な場合には個々の情報のカテゴリー・プロトタイプへの

統合が生じるために不一致情報の想起確率は著しく低下することになる。

また, Pryor, Ostrom, Dukerich, Mitchell と Herstein (1983) は, 複数の対象人物の行動を観察するような場合には, 記憶において対象人物とその行動が分離され, 特定の対象人物の行動であってもそれがその人物の帰属されたカテゴリーのプロトタイプと合致しない場合には, それと合致する他の対象人物の行動として想起されるという行動事例の分類自体におよぼすカテゴリー化の効果を報告している。

したがって, 観察すべき対象人物が多く, 記憶に把持すべき行動情報が増大し, かつ, 把持すべき期間が長期化するほど, 人は自らの情報処理能力の制約のもとに, 個々の行動情報のカテゴリー・プロトタイプへの統合を強め, 特定のカテゴリー・プロトタイプと合致する行動情報のみを想起するようになる。このような人の社会的情報の処理様式の特徴が, 特定の対象人物のそれぞれの行動次元での評価の分化度を低下させ (ハロー効果), 対象人物間での評価の分化度も低下させることになる (寛容性効果および尺度範囲の制限)。

本研究では, このようなカテゴリー化が, 人物評価や行動情報の想起にいかなる影響をおよぼしているかを実験的に検討する。

実験 1 カテゴリー化と人事評価

本実験では, ラベリング手続きによって事前に対象人物の所属するカテゴリーを設定したとき, 同一の行動情報のもとにいかに異なる評価

が形成されるかを検討する。

(1) 被験者, 40代から50代の男性80名。被験者はすべて中間管理層であり, 実験は人事考課のシュミレーション実習として行なわれた。

(2) ラベリング操作, 被験者に職場における部下数名を自由記述してもらい, この結果から彼らが部下の分類によく用いている, 職務遂行能力の高低と職場における協調性の有無という2次元を抽出した。この2次元の組み合わせにより, 図2に示すような4つのカテゴリーを設定し, 評価対象人物の行動情報を提示する前に, この人物の直属の上司による評価として, それぞれ20名ずつにこのカテゴリー内容を教示した (ラベリング手続き)。

(3) カテゴリー・プロトタイプの測定, 被験者全員に上記の4つのカテゴリーを提示し, 自分の部下あるいは同僚のなかでそれぞれのカテゴリーにもっともよくあてはまる人物を想起してもらい, この人物について8つの評価項目について評価を求めた (評価項目の内容については後述)。

(4) 行動情報の提示, 対象人物の5つの場面における行動を記述した小冊子を配布し, 黙読してもらった。なお, それぞれの場面は約1000字で記述され, 3つが職務遂行場面, 2つが対人的場面である。また, 記述の内容は, 「仕事はできるが, 職場での協調性には欠ける」というものである。

(5) 人事評価, 行動情報の提示後, 対象人物について8つの評価項目について7段階で評価を求めた。8つの評価項目は, 表1に示されるような代表的な考課項目であり, 尺度形式とし

図 2 カテゴリーのタイプと教示内容

能力 協調性	高	低
	有	仕事もできるし職場の人々との協調性もある (カテゴリーA)
無	仕事はできるが職場の人々との協調性に欠ける (カテゴリーB)	仕事もできないし職場の人々との協調性にも欠ける (カテゴリーD)

表1 カテゴリー・プロトタイプの評価

() 内は SD

評価項目	カテゴリーA (N=80)	カテゴリーB (N=80)	カテゴリーC (N=80)	カテゴリーD (N=80)	F 比 (3,76)
知識・技能	6.25 (0.64)	6.00 (0.79)	3.45 (0.94)	2.65 (1.04)	68.54 (P<.001)
理解・判断	6.10 (0.55)	5.20 (1.06)	3.65 (0.99)	2.65 (0.88)	40.50 (P<.001)
創意・工夫	6.20 (0.62)	4.95 (1.15)	3.35 (1.09)	2.60 (0.94)	34.01 (P<.001)
折衝・説得	5.80 (0.89)	3.20 (1.20)	4.00 (1.17)	2.45 (1.10)	7.47 (P<.001)
指導・統率	5.85 (1.04)	3.30 (1.17)	3.65 (1.14)	2.65 (1.14)	6.63 (P<.001)
協調性	6.15 (0.67)	3.10 (1.02)	5.35 (0.81)	2.50 (1.05)	13.00 (P<.001)
積極性	6.50 (0.69)	4.75 (1.02)	4.25 (1.07)	2.35 (1.04)	28.14 (P<.001)
責任感	6.25 (0.79)	4.35 (1.31)	4.05 (1.36)	2.20 (0.89)	20.20 (P<.001)

表2 対象人物の評価

() 内は SD

評価項目	カテゴリーA (N=20)	カテゴリーB (N=20)	カテゴリーC (N=20)	カテゴリーD (N=20)	F 比 (3,76)
知識・技能	6.20 (0.77)	6.40 (0.60)	6.00 (0.72)	5.85 (0.49)	2.24 (P<.10)
理解・判断	5.05 (1.43)	5.35 (0.93)	5.20 (1.36)	5.55 (0.94)	n. s.
創意・工夫	5.70 (1.13)	5.30 (0.92)	5.40 (0.99)	5.75 (0.64)	n. s.
折衝・説得	3.80 (1.28)	3.55 (1.00)	3.80 (1.36)	3.90 (1.21)	n. s.
指導・統率	4.40 (1.35)	4.55 (1.10)	4.30 (1.03)	4.65 (1.18)	n. s.
協調性	3.75 (1.33)	3.40 (1.23)	3.95 (1.15)	3.90 (1.37)	n. s.
積極性	5.95 (0.94)	6.20 (0.83)	5.60 (1.27)	5.75 (0.72)	n. s.
責任感	4.75 (1.25)	5.00 (1.12)	4.55 (1.32)	4.55 (1.15)	n. s.

ては図式評定法を用い、評価項目名を記したあとに、評価項目の具体的内容の説明を加え、それに7領域に区切られた線分を付したものである。また、線分のそれぞれの領域についてもそのレベルを明示した。

それぞれのカテゴリー・プロトタイプおよび対象人物についての評価結果は、1から7に等間隔で数値化する。

(6) 結果と考察

それぞれのカテゴリー・プロトタイプの評定値は表1に示す通りである。いずれの評価項目についても、カテゴリー間で大きな有意差が見出された。ここから、被験者は、それぞれのカテゴリーについて明確なプロトタイプ・イメージを有していることが確認できる。また、カテゴリーを構成する次元間でもその優位性に階

層が存在し、社会的側面にくらべ、知識・技能や理解・判断などの職務遂行に係わる能力的側面の方がより優位であることも推測される。

それぞれのカテゴリー化条件のもとでの、対象人物の評定値は表2に示す通りである。ここでは、プロトタイプ・イメージを構成するもっとも優位な次元と思われる知識・技能項目でのみカテゴリー化条件間での有意差が見い出された。ここから、被験者は、われわれが予測するほどには事前カテゴリー化による影響を受けておらず、カテゴリーを構成するもっとも優位な次元においてはプロトタイプへの同一化が生じているものの、それ以外ではむしろ提示された行動情報に即した評価を形成しているといえる。

これは、評価対象人物がひとりであり、かつ、行動情報提示直後に評価が行なわれたため、被験者は個々の行動情報の直接的な把持が可能であり、行動情報のカテゴリーへの統合を必要としなかったためと思われる。しかし、それでも、もっとも優位な次元においてプロトタイプへの同一化が生じているのは、事前カテゴリー化が行動情報の把持と想起のみでなく、その記銘に大きな影響をおよぼし、この次元に関連した行動情報の選択的な受容を生じさせていることを示している。これらの点については、評価対象人物を複数化し、かつ行動情報の把持期間をさらに長期化することで検討する必要がある。

実験2 カテゴリー化と行動情報の想起

本実験では、社会的カテゴリー化によって対象人物と評定者との集団成員性 (group membership) の異同が設定された場合に、対象人物の行動情報がどのように想起され、それによって対象人物についていかなる評価が形成されるかを検討する。

社会的カテゴリーの場合には、実験1の記述的カテゴリーのような構造化されたプロトタイ

プ・イメージは存在しないが、対象人物との心理的距離を規定し、対象人物が評定者と同一の集団に所属する場合には心理的距離は小さく、より好意的な (favorable) 行動情報ほど記銘あるいは想起されやすく、異なる集団に所属する場合には心理的距離は大きく、より非好意的な (unfavorable) 行動情報ほど記銘あるいは想起されやすくなる。したがって、社会的カテゴリー化は、評定者と対象人物との心理的距離を規定することによってやはり評定者の行動情報の処理を方向づけ、とくに評価における寛容性効果の発生に強く影響しているものと思われる。

(1) 被験者、男子大学生34名。実験は、被験者6名から8名に、評価対象人物となる実験協力者1名を加えて行なった。

(2) 社会的カテゴリーの設定、被験者と実験協力者を、3名から5名ずつのふたつのグループに分け、それぞれのグループごとに作業を行なうという状況を設定することで、評価対象人物 (実験協力者) とそれぞれの被験者グループとの集団成員性の異同を導入した。つまり、ふたつに分けられたグループのうち実験協力者も入っている方のグループの被験者にとっては対象人物は同じ集団の成員 (以後、内集団成員と呼ぶ) であり、他のグループの被験者にとっては対象人物は異なる集団の成員 (外集団成員と呼ぶ) ということになる。

(3) 行動情報の提示、社会的カテゴリーの設定後、評価対象人物を実験室から退室させ、残った被験者にこの対象人物の種々の場面での行動情報を24項目、実験者が口頭で提示する。情報項目は、課題達成場面での成功事例4項目、失敗事例4項目、対人的場面での協調行動事例4項目、攻撃行動事例4項目、中性的なダミー事例8項目からなり、それぞれが短文で示され、かつランダム順に2度読みあげられる。

(4) 行動情報の想起、行動情報を提示してから約5分後に、これらの情報項目の自由再生を行なった。自由再生では、情報項目の提示順序にかかわらず、被験者が想起する順に情報項目

表3 行動情報の想起得点

() は SD

課題達成場面			対人的場面		
対象人物の集団成員性	成功事例	失敗事例	対象人物の集団成員性	協調事例	攻撃事例
内集団成員 (N=16)	1.81 (1.29)	3.06 (1.89)	内集団成員 (N=16)	1.44 (1.73)	0.88 (1.45)
外集団成員 (N=18)	2.06 (2.15)	2.83 (2.14)	外集団成員 (N=18)	0.50 (0.90)	1.56 (2.22)

を記述することが求められるが、より早く想起される情報ほど認知構造への統合度は高く、対象人物についての評価をより強く規定することになる。自由再生の結果については、第1想起項目に4点、第2想起項目に3点、第3想起項目に2点、第4想起項目に1点、それ以外は0点と得点化し、それぞれの行動事例ごとに4項目の想起得点の総和を求めた。これは、情報項目の想起順位のそれぞれの行動事例へのかたよりをより敏感に検出しようとするためのものであり、たとえば、課題成功事例の4項目が最初に想起されるなら、この行動事例の想起得点は10点となり、他は0点となる。

(5) 対象人物の評価、行動情報の想起後に対象人物について8つの評価項目での評価を求めた。ただし、本実験では被験者は大学生であり、人事考課項目に不慣れなため、有能な一無能な、協力的な一攻撃的な、などの形容詞対からなるより一般的なSD尺度 (Semantic Differential Scale) を用い、これらについて5段階での評定を求めた。各項目の評定結果は、1から5へ等間隔で数値化し、8項目への評定値の平均を評価測度とした。

(6) 結果と考察

行動情報の自由再生の結果は表3に示す通りである。分散分析の結果では、課題達成場面では事例間の主効果のみが有意水準に達し ($F=4.81$, $df=1, 32$, $P<.05$), 成功事例にくらべ失敗事例の方がより想起されやすかった。しかし、われわれの予測する対象人物の集団成員性条件と行動事例の評価方向との交互作用効果は見い出されなかった。一方、対人的場面で

は、集団成員性と行動事例との交互作用効果のみが有意水準に達し ($F=5.42$, $df=1, 32$, $P<.05$), 対象人物が内集団成員の場合には好意的行動の方が想起されやすく、外集団成員の場合には非好意的行動の方が想起されやすかった。この結果は、われわれの仮説を支持するものである。

また、対象人物の評価は、表4に示す通りであり、対象人物が評定者と同じ集団に所属する場合には、ネガティブな行動の手掛りが多いにもかかわらず評価は比較的高く、対象人物が異なる集団に所属する場合には評価はきわめて厳しいものであった (両者の差は10%の有意水準に達する, $t=1.87$, $df=33$, $P<.10$)。

表4 対象人物の評価 () 内は SD

対象人物の集団成員性	評定値
内集団成員 (N=16)	2.88 (2.37)
外集団成員 (N=18)	0.94 (3.61)

ここから、社会的カテゴリー化は、評価における内集団成員への寛容性効果の出現を規定しているといえるが、行動情報全般の処理を方向づけるわけではなく、その影響は社会的場面における行動に限定されている。一方、実験1で見い出されたように、プロトタイプへの同一化は、とくに能力的側面において強くあらわれている。したがって、人事評価における評定誤差の出現は、いずれかひとつの情報処理の特質によるものではなく、人の社会的情報の処理様式全般の複合効果によるものであり、行動情報の処理全体についての理論的説明とそれに基づい

た人事評価過程のモデル化および実証的検討が必要とされよう。

(本研究における実験は吉田秀雄記念事業財団による助成研究の一部として行なわれたものである。ここに記して感謝の意を表する)

引用文献

- Amir, Y., Kovarsky, Y., & Sharan, S. Peer nominations as a predictor of multistage promotions in a ramified organization. *Journal of Applied Psychology*, 1970, 54, 462-469.
- Barrett, R. S., Taylor, E. K., Parker, J. W., & Martens, L. Rating scale content: I. Scale information and supervisory ratings. *Personnel Psychology*, 1958, 11, 333-346.
- Benardin, H. J. Behavioral expectation scales versus summated scales: A fairer comparison. *Journal of Applied Psychology*, 1977, 62, 422-427.
- Bernardin, H. J. Effects of rater training on leniency and halo errors in student ratings of instructors. *Journal of Applied Psychology*, 1978, 63, 301-308.
- Borman, W. C. Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. *Journal of Applied Psychology*, 1975, 60, 556-560.
- Borman, W. C., & Dunnette, M. D. Behavior-based versus trait-oriented performance ratings: An empirical study. *Journal of Applied Psychology*, 1975, 60, 561-565.
- Cantor, N., & Mischell, W. Prototypes in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, Vol. 12, Academic Press, 1979.
- Cohen, C. E. Person categories and social perception: Testing some boundaries of the processing effects of prior knowledge. *Journal of Personality and Social Psychology*, 1981, 40, 441-452.
- Cotton, J., & Stoltz, R. E. The general applicability of a scale for rating research productivity. *Journal of Applied Psychology*, 1960, 44, 276-277.
- DeCotiis, T. A. An analysis of the external validity and applied relevance of three rating formats. *Organizational Behavior and Human Performance*, 1977, 19, 247-266.
- Finn, R. H. Effects of some variations in rating scale characteristics on the means and reliabilities of ratings. *Educational and Psychological Measurement*, 1972, 32, 255-265.
- Friedman, B. A., & Cornelius, E. T., III. Effect of rater participation in scale construction on the psychometric characteristics of two rating scale formats. *Journal of Applied Psychology*, 1976, 61, 210-216.
- Gordon, M. E. The effect of the correctness of the behavior observed on the accuracy of ratings. *Organizational Behavior and Human Performance*, 1970, 5, 366-377.
- Hamner, W. C., Kim, J. S., Baird, L., & Bigoness, W. J. Race and sex as determinants of ratings by potential employers in a simulated work sampling task. *Journal of Applied Psychology*, 1974, 59, 705-711.
- Keaveny, T. J., & McGann, A. F. A comparison of behavioral expectation scales and graphic rating scales. *Journal of Applied Psychology*, 1975, 60, 695-703.
- Kingsbury, F. A. Analyzing ratings and training raters. *Journal of Personnel Research*, 1922, 1, 377-383.
- Klieger, W. A., & Mosel, J. N. The effect of opportunity to observe and rater status on the reliability of performance ratings. *Personnel Psychology*, 1953, 6, 57-64.
- Klores, M. S. Rater bias in forced-distribution ratings. *Personnel Psychology*, 1966, 19, 411-421.
- Landy, F. J., & Guion, R. M. Development of scales for the measurement of work motivation. *Organizational Behavior and Human Performance*, 1970, 5, 93-103.
- Latham, G. M., Wexley, K. N., & Pursell, E. D. Training managers to minimize rating errors in the observation of behavior. *Journal of Applied Psychology*, 1975, 60, 550-555.
- Mandell, M. M. Supervisory characteristics and ratings: A summary of recent research. *Personnel Psychology*, 1956, 32, 435-440.
- Newcomb, T. An experiment designed to test the validity of a rating technique. *Journal of Educational Psychology*, 1931, 22, 279-288.
- Nisbett, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 1977, 84, 231-259.
- Pryor, J. B., Ostrom, T. M., Dukerich, J. M., Mitchell, M. L., & Herstein, J. A. Preintegrative categorization of social information: The role of persons as organizing categories. *Journal of*

- Personality and Social Psychology*, 1983, 44, 923-932.
- Rothaus, P., Morton, R. B., & Hanson, P. G. Performance Appraisal and psychological distance. *Journal of Applied Psychology*, 1965, 49, 48-54.
- Saal, F. E., & Landy, F. J. The mixed standard rating scale: An evaluation. *Organizational Behavior and Human Performance*, 1977, 18, 19-35.
- 境 忠宏: 広告情報の認知的体制化とイメージ形成。吉田秀雄記念事業財団助成研究報告書, 1985.
- Schneier, C. E. Operational utility and psychometric characteristics of behavioral expectation scales: A cognitive reinterpretation. *Journal of Applied Psychology*, 1977, 62, 541-548.
- Scott, W. E., Jr., & Hamner, W. C. The influence of variations in performance profiles on the performance evaluation process: An examination of the validity of the criterion. *Organizational Behavior and Human Performance*, 1975, 14, 360-370.
- Stockford, L., & Bissell, H. W. Factors involved in establishing a merit-rating scale. *Personnel*, 1949, 26, 94-116.
- Taylor, E. K., Schneider, D. E., & Clay, H. C. Short forced-choice ratings work. *Personnel Psychology*, 1954, 7, 245-252.
- Thorndike, E. L. A constant error in psychological ratings. *Journal of Applied Psychology*, 1920, 4, 25-29.
- Vance, R. J., Kuhnert, K. W., & Farr, J. L. Interview judgments: Using external criteria to compare behavioral and graphic scale ratings. *Organizational Behavior and Human Performance*, 1978, 22, 279-294.
- Wyer, R. S., Jr., & Srull, T. K. Category accessibility: Some theoretical and empirical issues concerning the processing of social stimulus information. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario symposium on personality and social psychology*. Lawrence Erlbaum, 1980.

[さかい ただひろ 横浜国立大学経営学部助教授]