

主観的業績評価におけるバイアス： 分析的研究のレビューに基づく考察

森 光 高 大
木 村 太 一

西南学院大学商学論集
第68巻 第1・2合併号 抜刷
2021（令和3）年12月発行

主観的業績評価におけるバイアス： 分析的研究のレビューに基づく考察

森 光 高 大
木 村 太 一

1. はじめに

マネジメント・コントロール・システムは、従業員のモチベーションを喚起し、彼ら/彼女らから組織目標と統合的な努力を引き出すことを目的として設計される。これをうまく機能させるためには、管理可能かつ目的統合的な客観的業績評価指標を設定し、それを報酬・インセンティブと対応させることが基本になる (Lambert 2001)。しかし、多くの場合、すべての従業員が客観的な指標のみにもとづいて評価できるわけではなく、客観的評価による動機付けの限界もまた指摘されている。このとき、客観的指標を補うものとして注目されるのが、主観的評価である。実際、客観的に評価しづらいタスクに従事している従業員を評価する方法として、あるいは客観的指標だけでは捉えられない側面を評価する方法として、直属の上司等の主観的な判断にもとづく評価は広く用いられている。

一方で、主観的業績評価は、あくまで評価者の主観に基づいている以上、外部からの検証が困難となる。主観的業績評価においては、様々な評価の歪み、いわゆるバイアスが生じうることが、多くの先行研究によって指摘されている¹。こうしたバイアスには2つの主要な分類を提示すること

1 主観的業績評価において発生しうる他の問題として、評価者と被評価者が共謀して不当に高い評価を得るなどの「汚職」(collusion)の存在も指摘されている (Tirole 1986; Thiele 2013)。これは評価結果が第三者から立証不可能であるという性質を背景としたものである。不当に高い評価とそれに基づく報酬の高騰はプリンシパル (雇用者)の効用を害することとなるため、この「汚職」も主観的業績評価においては重要な論点であるといえるが、本稿の主眼はあくまで主観的業績評価における「バイアス」であるため、「汚職」は直接の検討対象とはしない。

ができる。1つが、評価者が従業員の評価を不当に甘くつけてしまう寛大性バイアス (leniency bias) である。もう1つが、評価者の付けるレーティングの分散が小さくなり、多くの従業員の評価が狭い範囲の中に収まってしまう中心性バイアス (centrality bias) である。これら2つのバイアスについては多くの研究において、その存在が認識されてきた。

自身の評価にバイアスが生じ、業務の成果が適切に評価されづらい状況であれば、従業員の動機づけが十分になされないことも懸念される。そのため、主観的評価において、このバイアスの原因や影響は、先行研究においても、実証や実験室実験、分析的研究など様々な研究方法を用いて盛んに議論されてきた。Moers (2005) は、業績評価指標の多様性、主観性と、評価におけるバイアスの関係について、オランダ企業の人事データを用いて実証を行っている。分析の結果、彼らは複数の客観的指標と主観的指標を使用する場合、評価結果が圧縮され、より寛大な評価になることを提示した。また、Bol (2011) は、オランダの金融機関の人事データを用いて、情報収集のためのコストおよび従業員と雇用主との間の強力な関係が、中心性バイアスと寛大性バイアスに影響を与えることを示した。彼女らは実証の結果、中心性バイアスは成果に負の影響を与えているが、寛大性バイアスは現在の成果に対して負の影響を与えるという有意な結果は出なかったことも提示している。Angelovski et al. (2016) は、実験室実験により、従業員の採用経緯を区分けし、評価者のバイアスについて調査した。実験の結果、強く推薦された従業員の評価には正のバイアスがかかり、推薦されなかった従業員の評価には負のバイアスがかかることを提示した。Kramer and Maas (2020) は、西ヨーロッパのビジネススクールの学生97名を被験者として、主観的評価に関する実験研究を行っている。彼らは、従業員の昇進に関するバランスト・スコアカード形式の評価レポートにおいて、評価者がより高い点数付ける傾向にあることを提示した。

主観的評価においてこうしたバイアスが生じる原因についても、様々な視点から多くの検討がなされている。代表例としては、評価者が特定の部下に対して思い入れを持ち、他の部下に比較して不当に高い評価を与え

る、「ひいき」(favoritism)を挙げることができる。こうした「ひいき」が生じる理由についても議論がなされており、人種、性別などを説明要因として研究が蓄積されている。人種について、Elvira and Town (2001)は、米国のデータを用いて、従業員の評価が、実際の生産性にかかわらず、人種によって変わりうることを示した。また、Price and Wolfers (2010)は、米国プロバスケットボールリーグであるNBA (National Basketball Association)において、審判と選手の人種の違いが審判の判定に影響を与えている可能性について示唆している。類似の研究として、Parsons et al. (2011)は米国のプロ野球リーグであるMLB (Major League Baseball)のデータを用いて、球審と投手の人種(民族集団)の違いが、球審の判定に影響を与えているという同様の可能性を指摘した。性別については、Goldin and Rouse (2000)が、米国のオーケストラの採用に関するデータを用いて、目隠しによるオーディションによれば、女性のミュージシャンが採用される可能性が上昇することを提示している²。

また、「ひいき」と類似のコンテキストで使われる概念として、部下に対する同情や、職場の雰囲気を崩したくないなどの理由から、低い評価を付けることに抵抗を感じたり、高い評価を付けてあげたいと感じる個人感情の存在も指摘されている。これらは「互恵性」(reciprocity)、「利他主義」(altruism)などの用語で説明されている。

実証研究、実験研究は、基本的にはバイアスの存在や影響を検証するものであるが、その発生メカニズムについて詳細に論じたものではない。こうしたメカニズム等について分析しているものが、分析的研究の数理モデルである。しかし、分析的方法による先行研究において、同様の用語が使われていたとしても、このバイアスの原因が特定のエージェントに対するものであったり、エージェント全体に対するものであったりなど、想定されている状況は必ずしも一様ではない。そのため、モデルの構造および

2 また、性別により主観的評価に差が出ないことを示した研究も存在する。Abrevaya and Hamermesh (2012)は論文査読のデータを用いて、著者と査読者の性別の違いは、査読結果に優位な影響を与えていないことを示している。

状況設定について相違の理解がなければ、議論に齟齬が生じることが懸念される。したがって、主観的業績評価に関する代表的な分析的研究を概観し、その類型を整理することは、今後の同分野の研究にも有用な知見を与えうる。

しかし、これまで、主観的業績評価に関して、分析的研究のモデルに立ち入りレビューした先行研究はほとんど見られない³。以上のような現状を踏まえ、本論文では、主観的業績評価に関する研究のうち、特に分析的研究の方法論を採ったものに焦点を当て、効用関数の形式やバイアスが生じる原因、過程について整理を行う。本稿では、これらのレビューを通じて、主観的業績評価における論点の明確化や、今後の研究余地を探ることを企図している。

2 本稿の分析のフレームワーク

本稿では、先行研究の分類を行うにあたり、効用関数の構造、特にバイアスの原因となる変数がどのように設定されているかという点に注目する。バイアスが特定のエージェントのみについてかかる場合と、エージェント全体に対して等しくかかる場合では、モデルの想定する状況は大きく異なる。そのため、このバイアスの対象は誰なのかという点を、本稿における分類の1つの観点とする。

さらに、こうした変数が、どのような原因で生じ、どのような影響を与えるかという点も、注視すべき論点である。そのため、バイアスの対象だけでなく、こうした変数がどのような設定のもとに生じ、分析結果にどのように作用しているのかという点も、分類のための重要な視点とする。

また、主観的評価のバイアスに関する先行研究には、Müller and Weinschenk (2015) など複数期間を想定しモデル分析を行ったものも存在するが、類似の枠組内で比較の合理性を維持するために、本稿では1期間の

3 北田(2015)は、主観的業績評価に関するレビューを行っている。当該論文は、会計研究エリアのジャーナルに的を絞り主観的業績評価の研究の潮流を整理したものであるが、分析的研究における効用関数の立て方やバイアスの取扱い方に関しては直接の検討対象とはしていない。

みについて扱ったものを検討対象とする。

3 主観的業績評価におけるバイアスを組み込んだ分析的研究

3-1 特定のエージェントの報酬に対するバイアス

先述したように、主観的評価においてバイアスが生じる原因の代表例として、評価者が特定の部下に対して思い入れを持ち、他の部下に比較して不当に高い評価を与える、「ひいき」(favoritism)を挙げることができる。これは、主観的評価においてバイアスが生じる主要な原因の1つとして認識され、多くの先行研究において、この「ひいき」を組み込んだモデルが分析されている。特定のエージェントが不当に高い（もしくは低い）報酬を得るように評価を行うというのが、「ひいき」のモデルの基本的な特徴であり、こうした研究の代表的なものとして、Prendergast and Topel (1996)を挙げることができる。まず当該論文のモデルを追いながら、主観的評価における「ひいき」について検討を行う⁴。

Prendergast and Topel (1996)のモデルの基本的な設定は下記のとおりである。まず、プリンシパルである企業が、従業員と監督者を一人ずつ雇用する。監督者は従業員の成果を私的に観察し、観察結果は立証不可能であるとする。この立証不可能性が評価の主観性を意味する。従業員の成果 y は下記の式で与えられる。

$$y = e + \alpha + \varepsilon_s \quad (1)$$

ここで、 e は従業員の努力投入量を意味する。 α は従業員の能力を意味し、これは平均0かつ分散 σ_α^2 の正規分布に従う($\alpha \sim N(0, \sigma_\alpha^2)$)。また、 ε_s は監督者の観察に際する誤差項であり、平均0かつ分散 σ_ε^2 の正規分布に従う

4 本稿における変数等の表記については、基本的に引用元の原文に従うものとする。そのため、異なる意味の変数でも部分的に表記が重複しているものがあるが、同一モデル内に混在するわけではないので特に混乱はきたさないと考える。また、明確な拡張関係により完全に意味を同じくするものについては、一部表記を統一している変数も存在する。

($\varepsilon_s \sim N(0, \sigma_\varepsilon^2)$)。さらに、 α と ε の間に相関はない。

従業員は下記のような指数的効用を有する。

$$v = -\exp\{-r[w - c(e)]\} \quad (2)$$

w は従業員の賃金であり、 r は絶対的リスク回避度である。また、従業員は e の努力を投入するために $c(e)$ のコストを要する。ここで、監督者の効用は次のようにモデル化されている。

$$v_s = w_s + \eta w \quad (3)$$

この η は正と負の両方の値を取りうると仮定され、特定の従業員に対する監督者の選好、つまり「ひいき」を示している。 η の事前分布は、 $\eta \sim N(0, \sigma_\eta^2)$ である。これは平均 0 かつ分散 σ_η^2 の正規分布に従うことを意味し、分散 σ_η^2 が大きくなるほど潜在的に大きなバイアスが生じうることを意味する。この監督者の主観的評価とは別に、プリンシパル（企業）も次のような形で独自の主観的評価を行っている。

$$y_f = e + \alpha + \varepsilon_f \quad (4)$$

企業はエージェントに対して特別な選好を有しておらず、監督者への報酬は次式のように示されている。

$$w_s = w_0 - \frac{1}{2} \lambda (y_s - y_f)^2 \quad (5)$$

このように、企業の評価と監督者の評価の乖離にペナルティをかけることによって、監督者が「ひいき」によって、従業員に不当に高い（もしくは低い）評価を付けることを制限している。また、このようにして行われた監督者と企業の評価にもとづき、従業員の報酬は下記のように決定される。

$$w = \tau_0 + \tau_s y_s + \tau_f y_f \quad (6)$$

Prendergast and Topel (1996) は、寛大性、中心性といった特定の形態のバイアスは論点としておらず、「ひいき」を表す変数 η が存在することによって、「ひいき」がない場合と比較して、評価が歪むという構造のモデルを分析している。上記のように、監督者の報酬は企業との評価の差に影響を受けるため、監督者は企業の私的情報を推測することとなり、バイアス $b(\eta, y, \lambda)$ を含んだ監督者の報告は次のようになる。

$$y_s = E(y_f | y, e^*) + b(\eta, y, \lambda) \quad (7)$$

ここでの e^* は均衡でエージェントが投入する最適な努力水準であり、監督者の推測する企業の私的情報は、次式で示される。

$$E(y_t | y, e^*) = \theta y + (1 + \theta) e^* \quad (8)$$

ここで、 $\theta = \sigma_a^2 / (\sigma_a^2 + \sigma_s^2)$ である。(6)式に(7)式と(8)式を代入すると、従業員の報酬は次式のようになる。

$$w = \tilde{\tau}_0 + \tilde{\tau}_s y + \tau_f y_f \quad (9)$$

この式の $\tilde{\tau}_0$ と $\tilde{\tau}_s$ はそれぞれ、 $\tilde{\tau}_0 = \tau_0 + \tau_s(1 - \theta)e^* + \tau_s b$ 、 $\tilde{\tau}_s = \tau_s \theta$ である。これらの式に基づき、 η を主要因とする監督者のバイアスは次式により選択される。

$$\max_b \eta \tau_s [\theta y + (1 - \theta) e^* + b] - \frac{\lambda}{2} b^2 \quad (10)$$

これにより、下記の式が得られる。

$$b = \frac{\eta\tau_s}{\lambda} \quad (11)$$

この式から、監督者のバイアスは、エージェントに対する選好、つまり「ひいき」の度合である η と監督者のレポートの影響力 τ_s について増加的であることがわかる。つまり、監督者は、自身のレポートが従業員の報酬により強いインパクトを持つ場合に、より強いバイアスを生じさせる。また、バイアスは、監督者と企業の評価の乖離に対するペナルティの度合い λ については減少的であり、この λ を Prendergast and Topel (1996) は「ひいき」の価格 (The price of favoritism) と表現している。

Prendergast (2002) は「ひいき」について、Prendergast and Topel (1996) を拡張し、主観的に評価を行う監督者のもとでの最適な成果報酬や、リスクとインセンティブのトレードオフを分析している。プリンシパル（企業）が監督者とエージェントを雇用するという設定は共通しているが、企業と監督者がそれぞれ独自に主観的評価を行うのではなく、あくまで従業員の評価を直接的に行うのは監督者であるという設定を置いている。これは、現場の評価はトップマネジメントではなく現場の監督者が行うという、より現実的な設定であると解釈できる。主観的な評価のベースとなるシグナル y は次式の通りに表現されている。

$$y = e + \alpha + \varepsilon \quad (12)$$

これを監督者が観察し、主観的評価の結果として、下記のレポートを提出する。

$$y_s = y + b(y) \quad (13)$$

この b は、Prendergast and Topel (1996) と同様に、監督者が評価を歪めた度合であり、主観的評価にある程度の恣意性が介在することを仮定し

ている。ここでも監督者が評価を歪ませる原因となるのが、「ひいき」の存在である。Prendergast (2002) は、監督者は特定の従業員に対する選好を有すると仮定し、効用を下記のように設定している。

$$v_s = w_s + \eta w - \frac{b^2}{2} \quad (14)$$

この監督者の効用の右辺第2項までは Prendergast and Topel (1996) と同じであるが、恣意的な歪みを制約する条件に違いがある。ここでの b は上記の通り、監督者が評価を歪ませた度合であるが、真の成果と乖離した評価をつけるためには、それを取り繕う (covering up) ためのコストがかかり、また、あまりに不可解な評価を付けた場合は、監督者に解雇の可能性も生じる。こうしたコストをここでは $b^2/2$ として表現している。

Prendergast (2002) は変数 η について「利他主義 (altruism)」という用語で説明し、「ひいき」と「利他主義」をほぼ同義で使用している。しかし、この η は定数ではなく、エージェントによって異なるうえ、エージェントの報酬に掛かる変数という設定上、「特定のエージェントの報酬に対する」利他主義、つまり「ひいき」の意味合いが大きいといえる。

3-2 不均衡回避によるバイアス

上述した Prendergast and Topel (1996)、Prendergast (2002) のようなモデルは、プリンシパルもしくは評価者がエージェントの報酬を考慮することによってバイアスが生じる点に特徴があったが、報酬ではなくエージェントの効用それ自体を考慮するモデルも存在する。これらは、エージェントに不当に高い（もしくは低い）報酬を与えるというのではなく、エージェントの効用が高くなることによって、プリンシパルの効用も高くなるという性質のものである。先述したように、「ひいき」のモデルにおいても「利他主義」という用語は使われていたが、こうした設定は、より明白に「互惠性」や「利他主義」を示していると捉えることができる。

Grund and Przemeczek (2012) は、プリンシパルの効用関数の一部にエー

エージェントの効用がそのまま包含されるモデルを用いて、エージェントの不均衡回避によってバイアスが生じることを分析している。当該論文は、不均衡回避をモデル化するために、企業が一人の監督者と2人のエージェント i, j を雇用するという状況を想定している。

監督者はエージェントの真の成果 $p_i (>0)$ を観察し、それに基づき、立証不可能な成果レポート r_i を提示する。このレポートにより、エージェントの報酬が $W_i = w + br_i$ という線形契約で決定される。 w は固定報酬であり、 b は評価レポートにかかるボーナスレートである。ここで、双方のエージェントはともに、自身の評価（報酬）がもう一方より低い場合、嫉妬のような感情をもち、これが効用関数を害することとなる。これが不均衡回避であり、エージェント i については、 $(r_j - r_i)$ で示される。これを踏まえたうえで、エージェント i の効用は下記のように示される。

$$\begin{aligned} U_i(r_i, r_j) &= W_i - \alpha (\max\{W_j - W_i, 0\}) - \beta (\max\{W_i - W_j, 0\}) \\ &= w - br_i - \alpha (\max\{b(r_j - r_i), 0\}) - \beta (\max\{b(r_i - r_j), 0\}) \\ & \quad i \neq j \quad \text{and} \quad 0 \leq \beta < 1, 0 \leq \alpha \end{aligned} \quad (15)$$

ここで、 α は不均衡回避にかかる係数であるが、 β は自身ともう一方の評価に差が発生した場合に、もう一方を思いやる感情や罪悪感のようなものを示している。特殊なケースであるが、もし $\alpha = \beta = 0$ であれば、評価は気にせず、自身の報酬のみに関心を持つことを意味する。上式で Grund and Przemeczek (2012) は $\beta < 1$ を想定しているが、これはエージェントがあくまで他者よりも自身の効用を重視するという状況をモデル化するためである。

監督者は自身のレポートの正確に関心を持ち、自身の観察した評価とレポートが乖離すると $v(r_i - p_i)$ のコストが生じるものと仮定する。これは監督者の一定程度の倫理的行動のモデル化として説明されている。さらに、監督者はエージェントとの衝突を避けるため、職場の雰囲気が悪くしたくないなどの理由から、エージェントの効用にも配慮している。この配慮の

強さは μ で表され、監督者の効用関数は下記のように示されている。

$$U_s(r_i, r_j | p_i, p_j) = \mu (w + br_i - \alpha(\max\{b(r_j - r_i), 0\})) - \beta(\max\{b(r_i - r_j), 0\}) + w + br_j \\ - \alpha(\max\{b(r_j - r_i), 0\}) - \beta(\max\{b(r_i - r_j), 0\}) - v(r_i - p_i)^2 - v(r_j - p_j)^2 \quad (16)$$

ただし、 $\mu \geq 0$ である。

Grund and Przemeczek (2012) は、本モデルの分析の結果から、双方の成果が異なる場合 ($p_i > p_j$)、不衡平回避の度合いと成果の差が十分に大きければ、監督者は双方で異なりはするが、不衡平回避に配慮したレポートを提示し、十分に大きくなければ双方で同一のレポートを提示することを示した⁵。異なる評価を付ける場合も、不衡平回避を考慮し、報酬の差を埋めようとする評価戦略がとられており、これが中心性バイアスを意味する。また、異なる評価が付けられる場合、真の成果より高いレポートが提示されることとなり、双方の評価を調整する場合も、優れたエージェントの下方修正と劣ったエージェントの上方修正の合計は常に正となる。この状況が、寛大性バイアスの発生を意味する。また、同一の評価が付けられる場合でも基本的には同様で、低い成果のエージェントの評価も寛大性により引き上げられることとなるが、その度合いは不衡平回避の程度に依存する。

監督者の、エージェントの効用への考慮については、先述した Prendergast and Topel (1996)、Prendergast (2002) で示されている「エージェントの報酬への配慮」という要因と大きくは違わない。しかし、エージェント間の不衡平回避がこれら中心性、寛大性バイアスの原因となりうるという結果を示したことは、当該研究の特徴であるといえる。

3-3 効用関数がプリンシパルとエージェントの間で相互に入り込むモデル

Dur and Tichem (2015) は、経営管理者（プリンシパル）と従業員（エー

5 言い換えれば、 $p_i - p_j > \mu b(\alpha + \beta)/v$ であれば、監督者は $r_i^{diff} = \mu b(1 - \alpha - \beta)/2v$ および $r_j^{diff} = \mu b(1 + \alpha + \beta)/2v + p_j$ を提示する。 $p_i - p_j \leq \mu b(\alpha + \beta)/v$ であれば、 $r_i^{diff} = r_j^{diff} = \mu b/2v + (p_i + p_j)/2$ を提示する。

ジェント)の間の利他主義が、最適契約にどのような影響を与えるかを分析している。当該論文は、本文中では一貫して利他主義 (altruism) という用語を用いているが、効用関数の設定上、プリンシパルおよびエージェントの効用の一部に互いの効用がそのまま入るという構造になっているため、特に「互恵性 (reciprocity)」のモデルであると解釈できる。Dur and Tichem (2015) では、1人のプリンシパルが1人のエージェントを雇用する状況が設定されており、両者がともにリスク中立的であると仮定されている。まず、プリンシパルの効用は次のように設定されている。

$$\Pi = \pi + \gamma U \quad (17)$$

ここで、 π はエージェントの努力 e によって得られる成果 $q(e)$ から、エージェントへの支払い w を控除したものを意味し、 $\pi \equiv q(e) - w$ と表される。加えて、 U はエージェントの効用であり、 γ はプリンシパルのエージェントに対する選好や利他性を意味する。この部分だけを見ると、Prendergast and Topel (1996) のような「ひいき」のモデルと類似しているように感じるが、Dur and Tichem (2015) は次式で示されるエージェントの効用関数に特徴を見出すことができる。

$$U = u + \alpha \Pi \quad (18)$$

ここで、 u は支払われた報酬から努力コストを控除したエージェントの私的効用を意味し、 $u \equiv w - \theta e$ と表されている。重要となるのが次の $\alpha \Pi$ である。 Π は既述のように、プリンシパルの効用であり、 α はエージェントのプリンシパルに対する利他性を意味する。つまり、プリンシパル、エージェント双方の効用関数に互いの効用が含まれているのである。こうした双方向の利他性がまさに「互恵性」を直接的に表していると解釈できる。

また、双方の利他性である α および γ はこの関係に特定のな変数であり、ともに双方から観察可能であると想定されている。さらに、どちらも

自身の効用以上には他者の効用を重視しておらず、 $\alpha, \gamma \in (-1, 1)$ と設定されている。これら設定および、双方の効用が互いの効用に入り込むという構造から、プリンシパルおよびエージェントの効用関数は次のように表現される。

$$U = \frac{1}{1 - \alpha\gamma} (u + \alpha\Pi) \quad (19)$$

$$\Pi = \frac{1}{1 - \alpha\gamma} (\pi + \gamma U) \quad (20)$$

これら効用関数のもとで、エージェントは努力水準 e を低水準か高水準のバイナリで選択する。つまり、 $e \in \{0, 1\}$ と表記される。エージェントが高水準の努力を選択するためのコストは θ であり、実現する成果量 q は $q=H$ 、低水準の場合は 0 であり、実現する成果量 q は $q=L$ と設定されている。この努力と成果の評価が主観にもとづいて行われ、第3者からの立証が不可能という設定が置かれている。

分析の結果、Dur and Tichem (2015) は、利他主義は解雇の脅威による圧力を弱めさせ、ボーナスによるモチベーションを強くさせることを提示した。この2つのメカニズムは、特に、より強い利他主義はより高いボーナスを導出し、利他主義が弱まると、生産性とプレイヤーの効用を増加させることを含意している。

3-4 エージェント全体に対する「利他主義」

既述のように、バイアスが生じうる主要な原因の1つとして、多くの先行研究においては、「利他主義」という概念が説明されている。これは、評価者が様々な理由から、エージェントに高い評価をつけたがる、もしくは、低い評価を付けることに抵抗を感じるというものである。本稿ですでに検討した Prendergast and Topel (1996)などは、この「利他主義」を、特定のエージェントに対して掛かる変数としてモデル化し、「ひいき」という用語でも説明していた。

しかし、分析する状況によっては、「利他主義」を、すべてのエージェントに対して同様に作用する変数と設定している研究も存在する。このような設定における「利他主義」は、「ひいき」の文脈で扱われていたものとは大きく性質が異なり、評価者自身の本来の特性のような意味合いで捉えることができる。本項では、こうしたエージェント全体に対して等しく「利他主義」が発揮されるモデルを検討する。

Golman and Bhatia (2012) は、評価者の効用にエージェント全体に対する利他主義を組み込み、主観的業績評価における評価戦略について分析しているが、その過程のなかで、中心性バイアスと寛大性バイアスが発生する構造を提示している。

プリンシパル（経営管理者）は、評価者と、あるエージェント（従業員）と雇用契約を結ぶ。従業員は真の能力 x_i を有し、これは実数とする。プリンシパルは、この x_i を直接的に知ることはできないが、どのように分布しているかは知っているものとする。ここでは、平均 x で、分散 θ^2 の正規分布に従うと仮定する ($x_i \sim N(x, \theta^2)$)。この分布は事前知識であるが、その後、評価者はエージェントの成果に関するシグナル y_i を受け取る。このシグナルは、エージェントの真の能力 x_i に依存し、それに σ^2 の不確実性が加わったものとして、 $y_i \sim N(x_i, \sigma^2)$ とする。つまり、 y_i は平均 x_i 、分散 σ^2 の正規分布に従うとする。

エージェントの成果に関するシグナル y_i を受け取ったのち、評価者は主観的な評価である z_i (これも実数) を付ける。ここで、評価者の効用は次のように仮定される。

$$U_M(z_i) = \begin{cases} -\lambda(x_i - z_i) & \text{if } z_i < x_i \\ -(z_i - x_i) & \text{otherwise} \end{cases} \quad (21)$$

ここで、 λ は 1 より大きいことが仮定され、「評価者が、主観的に本来より高い評価を付ける場合と比較し、低い評価を付けることに相対的に抵抗を感じる」という状況がモデル化されている。この λ が、バイアスが生

じる直接の原因となる⁶。この状況は、特に評価者が残余請求者 (residual claimant) でない場合に発生し、その原因は従業員に対する共感や同情、職場の雰囲気などの要因で説明されている⁷。

分析の結果、プリンシパルの主観的な評価戦略は $\zeta(y_i)$ という関数形で表現され、その詳細は下記の通りとなる。

$$\zeta(y_i) = \bar{x} + (y_i - \bar{x}) \frac{\theta^2}{\sigma^2 + \theta^2} + \sqrt{2 \frac{\sigma^2 \theta^2}{\sigma^2 + \theta^2}} \operatorname{erf}^{-1} \left(\frac{\lambda - 1}{\lambda + 1} \right) \quad (22)$$

ただし、 $\operatorname{erf}(\cdot)$ は誤差関数である。

上式の右辺の成果の第2項 $\theta^2/(\sigma^2 + \theta^2)$ が中心性バイアスとして説明されている。シグナル y_i の分散は $\sigma^2 + \theta^2$ であり、エージェントの能力 x_i の分散は θ^2 である。そのため、シグナルの方がノイズが大きく、これらの分散のバランスを取るために、プリンシパルの評価戦略は平均に向かって圧縮されることになる。

また、右辺の最終項が寛大性バイアスとして説明されている。これは、 σ, θ, λ によって増加することとなる。もし、プリンシパルが低評価と高評価を無差別に考える ($\lambda = 1$) 場合、この項は0となるが、低評価を付けることに抵抗を感じれば感じる (λ が大きくなる) ほど評価は上昇することになる。

3-5 タスクの選好によるバイアス

これまでに検討した研究の多くは、「ひいき」や「利他主義」、低評価を付けることに抵抗を感じるなど、人物に対する感情をバイアスの源泉として捉えたものであった。しかし、Delfgaauw and Souverijn (2016) はこう

6 ここで、仮に $\lambda = 1$ であれば、プリンシパルは高い評価を付けることと、低い評価を付けることは無差別であるため、バイアスは生じない。

7 ただし、評価指標の不確実性が0 ($\sigma = 0$) であれば、実質的にプリンシパルはエージェントの能力を直接観察できるため、バイアスは生じず、 $x_i = y_i = z_i$ が最適となる (Golman and Bhatia 2012)。

した人物に対するものではなく、業務（タスク）の選好を、バイアスを生じさせる変数として組み込み分析を行っている。Delfgauw and Souverijn (2016) はプリンシパルが1人のエージェントと1人の監督者を雇用して、2つの業務(1, 2)に従事させるというマルチタスクの状況を想定している。エージェントと監督者の外部効用は0であり、双方ともに有限責任制約に守られているため、エージェントの報酬 w_A および監督者の報酬 w_S は0以上でなければならない。この状況下で、プリンシパルの効用関数は下記のように設定されている。

$$U_p = e_1 + e_2 - w_A - w_S \quad (23)$$

ここで e_1, e_2 はそれぞれの業務に対し、エージェントが投入した努力を意味する。この効用関数から明らかなように、プリンシパルにとってはどちらの業務も等しく価値を持つ。また、エージェントの効用は、受け取った報酬から努力コストを控除したものとして、次のように示されている。

$$U_A = w_A - \frac{1}{2}(e_1)^2 + \frac{1}{2}(e_2)^2 \quad (24)$$

ここで、プリンシパルはエージェントの努力を直接的に観察はできないが、監督者、エージェントがそれぞれに特定のタスクに対して選好を有すると仮定している。また、これはあくまでタスクに対する選好であり、特定のエージェントなどの人物に対する選好ではない。

プリンシパルはエージェントの努力を直接観察することはできず、努力に依存する立証可能な業績評価指標 m に頼ることとなる。しかし、この指標は、下記の式で表現されるように、特定のタスクに評価の比重が偏るバイアスがかかっている。

$$m(e_1, e_2) = \phi e_1 + (1 - \phi) e_2 \quad (25)$$

ここで、 $\phi \in [0, 1]$ であり、これが $1/2$ でなければ、どちらかのタスクに選好が偏ることとなる。また、ここで、この評価指標 m は完全に信頼はできないという仮定も置かれている。これは、エージェントが努力を投入後、 $(1-q)$ の確率で、コストを伴わず m を自身の望む水準に操作できると仮定することによりモデル化されている。

監督者の業務は、エージェントをモニターすることのみであり、 p の確率で、エージェントにより投入された努力 e_1 、 e_2 を観察することができる。監督者が観察した情報自体はソフトであり立証不可能であるが、この観察にもとづき、レポート r を作成する。このレポートがエージェントの報酬の基礎となるが、これはチープトークであり、どの水準でも作成することができる。監督者はエージェントの操作については直接的に観察することはできないが、努力水準からある程度推測することはできる。

監督者もエージェントの特定のタスクについて選好を持つ⁸ものとし、監督者の効用は次式で表記される。

$$U_s = w_s + \eta e_1 + (1 + \eta) e_2 \quad (26)$$

また、 $\eta \in [0, 1]$ であり、これも $1/2$ でなければ、監督者の選好はどちらかのタスクに偏ることとなる。このような構造のもとで、エージェントの報酬 w_A は、測定された成果 m と監督者のレポート r にもとづいて決定され、客観的な成果報酬 $b(m)$ と監督者の主観を含む $c(m, r)$ によって構成されるため、次のように表記される。

$$w_A(m, r) = b(m) + c(m, r) \quad (27)$$

分析の結果、彼らは、監督者が主観的評価を通じて、部下（エージェント）を、組織（プリンシパル）の目的と整合しない活動へ誘導しうること

8 この選好の原因としては、キャリアコンサーン、専門職としての責務、タスクの成果に対する内在的な考慮などが挙げられている (Delfgaauw and Souverijn 2016)。

を提示した。この監督者の行動は、立証可能な成果をうまく利用することにより緩和できるが、監督者の選好がより歪んでいる（つまり $\eta > \phi$ ）場合、完全に排除することはできない。こうした結果から、プリンシパルは、監督者の選定について、能力は低いがタスクの選好が歪んでいない者と、能力は高いが選好の歪みも大きい者というトレードオフに直面することが示されている。

Delfgaauw and Souverijn (2016) においては、エージェントおよび監督者のタスクの選好である、 ϕ, η がバイアスの原因となる。彼らも中心性バイアス、寛大性バイアスという用語には言及しておらず、論点は、エージェントと監督者がプリンシパルと整合しないタスクへの選好を持つために、プリンシパルはそれを考慮したうえで契約を設計しなければならないというものである。これまで見てきたように、主観的業績評価においては、多くの先行研究が、人物の評価やそれに伴う報酬や効用に対してバイアスがかかることを想定しているが、特定のタスクへの選好がバイアスの原因となるという分析を行ったことは、当該研究の大きな特徴であるといえる。

4 考察

本稿では、ここまで、主観的業績評価におけるバイアスについて、代表的な先行研究の、特に効用関数に注目し検討を行ってきた。主観的業績評価に関しては、実証研究、理論研究を含む多くの先行研究において、評価者にバイアスが生じることが指摘されている。しかし、バイアスの原因および、発生過程については、先行研究において多様なモデル化がなされている。そのため、同様、もしくは類似の用語が使われていても、モデルによって想定している状況およびバイアスの概念は少し異なる。

まず、これまでの検討でも明らかなように、バイアスの原因となるプリンシパルもしくは監督者の心因的な変数を、同じ「利他主義 (altruism)」という用語で説明している、それが特定のエージェントに対するものなのか、それとも全てのエージェントに対し等しく影響を与えるものなのかによって、意味合いは大きく変わってくる。Prendergast (2002) と Golman

and Bhatia (2012) は、バイアスの原因となる変数を、同様に「利他主義 (altruism)」という用語で説明している。しかし、Prendergast (2002) は Prendergast and Topel (1996) と同様に、特定のエージェントに対する「利他主義」、つまり「ひいき (favoritism)」の概念を想定しており、Golman and Bhatia (2012) は、エージェント全体に等しく影響する、評価者の本来的な特性のようなものとして「利他主義」を想定している。

また、Dur and Tichem (2015) も一貫して「利他主義」という用語を使っているが、彼らのモデルでは、プリンシパルとエージェントの効用関数に互いの効用が組み込まれ、それらに対して相互に「利他主義」が影響している。そのため、Dur and Tichem (2015) で用いられている「利他主義」は、他のものと異なり、より「互惠的」な特徴を有しているといえる。さらに、Dur and Tichem (2015) は、「互惠性」による評価の歪みを分析に含めてはいるが、本文中で一度もバイアスという用語では説明していないという特徴もある。このように、同じ「利他主義」という用語でも、設定により、その意味するものには相違があることがわかる。

次に、主観的に評価を行う評価者がプリンシパルなのか、それとも、あくまでプリンシパルに雇用された監督者なのかという点にも注意が必要である。Prendergast and Topel (1996)、Grund and Przemeczek (2012)、Golman and Bhatia (2012)、Delfgaauwand Souverijn (2016) は、プリンシパルが監督者とエージェントをそれぞれ雇用するという設定を行っているが、Dur and Tichem (2015) は、プリンシパル自身が評価者となっている。雇用された監督者は、エージェントと同じく、自身の評価にもとづきプリンシパルから受け取る報酬によって効用が決定されるが、評価者自身がプリンシパルである場合は、エージェントに努力を促すことが自身の効用につながる。そのため、これは、監督者の効用関数および評価戦略に直接的に関連する設定であり、主観的業績評価のモデルを参照するにあたり、認識しておくべき重要な着眼点となる。

これらに加え、バイアスの原因とそれらが生じる過程も注目に値する。本稿の冒頭でも述べたように、多くの先行研究は、主観的業績評価におけ

るバイアスを、中心性バイアス、寛大性バイアスという2つの概念で論じており、Grund and Przemeczek (2012)、Golman and Bhatia (2012) も、分析の結果としてこれら2つのバイアスが生じることを提示している。しかし、その原因、発生過程の解釈は異なっている。Grund and Przemeczek (2012) は、エージェント間の不均衡回避と評価者の「特定のエージェントに対する利他主義」をこれらバイアスの原因としており、Golman and Bhatia (2012) は、評価者の低評価を付けることに対する抵抗（「全てのエージェントに等しく影響する利他主義」）が原因となっている。分析的研究の効用関数まで遡り、こうしたバイアスの原因や発生過程の多様性および詳細を認識することは、実証的な研究を行う際などに有用な視座となるといえよう。

本稿のレビュー結果をまとめると、下記の表のように整理することができる。

	バイアスの原因	バイアスの対象	評価者の立場	主な分析結果とバイアスの影響
Prendergast and Topel (1996) Prendergast (2002)	「ひいき」の変数である η	特定のエージェント	プリンシパルが評価者を雇用	監督者のバイアスは、「ひいき」の度合と監督者のレポートの影響力について増加する。つまり、監督者は、自身のレポートが従業員の報酬により強いインパクトを持つ場合に、より強いバイアスを生じさせる。また、バイアスは、監督者と企業の評価の乖離に対するペナルティの度合いについては減少する。
Grund and Przemeczek (2012)	2人のエージェント間の評価の差である $(r_1 - r_2)$ と、監督者がエージェントの効用に配慮する μ	$(r_1 - r_2)$ はエージェントによって異なるが、 μ は2人のエージェントに等しくかかる	プリンシパルが評価者を雇用	双方の成果が異なる場合、不公平回避の度合いと成果の差が十分に大きい場合、監督者は双方で異なりはするが、不公平回避に配慮したレポートを、十分に大きくなければ双方で同一のレポートを提示する。 異なる評価を付ける場合、不公平回避を考慮し、報酬の差を埋めようとする評価戦略がとられる（中心性バイアス）。異なる評価が付けられる場合、真の成果より高いレポートが提示される（寛大性バイアス）。同一の評価が付けられる場合も低成果エージェントの評価も寛大性により引き上げられることとなるが、その程度は不公平回避の程度に依存。
Dur and Tischem (2015)	プリンシパルとエージェントが双方の効用に配慮し α と γ	特定のエージェントおよびプリンシパル	プリンシパル自身が評価者	利他主義は解雇の脅威による圧力を弱めさせ、ボーナスによるモチベーションを強くさせることを提示。
Golman and Bathia (2012)	プリンシパル（評価者）が低評価を付けることに抵抗を感じるという λ	エージェント全体	プリンシパルが評価者を雇用	エージェントの能力と成果のシグナルでは、成果のシグナルのほうが分散（ノイズ）が大きいため、これらのバランスをとるために、プリンシパルの評価戦略は平均に向かって圧縮される（中心性バイアス）。 プリンシパルが低評価を付けることに少しでも抵抗を感じれば、評価戦略は真の成果よりも上昇することとなる（寛大性バイアス）。
Delfgauuw and Souverijn (2016)	監督者およびエージェントの業務に対する選好 ϕ と η	特定のタスク	プリンシパルが評価者を雇用	監督者が主観的評価を通じて、エージェントを、プリンシパルの目的と整合しない活動へ誘導しうる。この監督者の行動は、立証可能な成果をうまく利用することにより緩和できるが、監督者の選好がより歪んでいる場合、完全に排除することはできない。 プリンシパルは、能力は低いタスクの選好が歪んでいない監督者と、能力が高いが選好の歪みも大きい監督者というトレードオフに直面する。

表1 本稿のレビューの整理

5 結論

本稿では、分析的研究の方法論が採られた代表的な研究についてレビューを行い、主観的業績評価において発生するバイアスについて整理、検討を行った。本稿の考察の結果、「利他主義」のように、同様の用語を使用しているも、モデルによって前提としている思考は異なっていることがわかる。この「利他主義」が特定のエージェントに対して生じているのか、すべてのエージェントに対して等しく生じているのかによって、その意味するところや分析結果への影響は大きく異なる。

また、分析結果に中心性、寛大性といった同様のバイアスが提示されていても、その原因や発生過程の解釈はモデルの設定により多様である。加えて、プリンシパル自身が評価者なのか、それともあくまでプリンシパルに雇用された監督者なのかという点も、評価戦略に直結する設定であり、重要な着眼点として指摘できる。

上述した内容を踏まえ、分析的研究においてモデルの拡張を試みる際も、実証研究のための仮説設定を行う際も、先行研究の設定および分析結果を参照する際には、使用されている用語のみで判断するのではなく、こうした変数がどのように使われているかを正しく認識する必要性が指摘できる。これらの相違は、先行研究の効用関数の設定を検討することにより峻別可能であり、この整理や明確化が本稿の主要な目的であったといえる。

【謝辞】

本研究はJSPS科研費 JP18K12897、JP21K01803、JP21K13406の助成を受けたものです。この場を借りて御礼申し上げます。

参考文献

- Abrevaya, Jason and Daniel S Hamermesh (2012) “Charity and favoritism in the field: Are female economists nicer (to each other)?” *Review of Economics and Statistics*, Vol. 94, No. 1, pp. 202–207.

- Angelovski, Andrej, Jordi Brandts, and Carles Sola (2016) “Hiring and escalation bias in subjective performance evaluations: A laboratory experiment,” *Journal of Economic Behavior & Organization*, Vol. 121, pp. 114–129.
- Bol, Jasmijn C (2011) “The determinants and performance effects of managers’ performance evaluation biases,” *The Accounting Review*, Vol. 86, No. 5, pp. 1549–1575.
- Delfgaauw, Josse and Michiel Souverijn (2016) “Biased supervision,” *Journal of Economic Behavior & Organization*, Vol. 130, pp. 107–125.
- Dur, Robert and Jan Tichem (2015) “Altruism and relational incentives in the workplace,” *Journal of Economics & Management Strategy*, Vol. 24, No. 3, pp. 485–500.
- Elvira, Marta and Robert Town (2001) “The effects of race and worker productivity on performance evaluations,” *Industrial Relations: A Journal of Economy and Society*, Vol. 40, No. 4, pp. 571–590.
- Goldin, Claudia and Cecilia Rouse (2000) “Orchestrating impartiality: The impact of “blind” auditions on female musicians,” *American Economic Review*, Vol. 90, No. 4, pp. 715–741.
- Golman, Russell and Sudeep Bhatia (2012) “Performance evaluation inflation and compression,” *Accounting, Organizations and Society*, Vol. 37, No. 8, pp. 534–543.
- Grund, Christian and Judith Przemeczek (2012) “Subjective performance appraisal and inequality aversion,” *Applied Economics*, Vol. 44, No. 17, pp. 2149–2155.
- Kramer, Stephan and Victor S. Maas (2020) “Selective attention as a determinant of escalation bias in subjective performance evaluation judgments,” *Behavioral Research in Accounting*, Vol. 32, No. 1, pp. 87–100.
- Moers, Frank (2005) “Discretion and bias in performance evaluation: the impact of diversity and subjectivity,” *Accounting, Organizations and*

Society, Vol. 30, No. 1, pp. 67–80.

- Müller, Daniel and Philipp Weinschenk (2015) “Rater bias and incentive provision,” *Journal of Economics & Management Strategy*, Vol. 24, No. 4, 833-862.
- Parsons, Christopher A, Johan Sulaeman, Michael C Yates, and Daniel S Hamermesh (2011) “Strike three: Discrimination, incentives, and evaluation,” *American Economic Review*, Vol. 101, No. 4, pp. 1410–1435.
- Prendergast, Canice (2002) “Uncertainty and incentives,” *Journal of Labor Economics*, Vol. 20, No. S2, pp. S115–S137.
- Prendergast, Canice and Robert H Topel (1996) “Favoritism in organizations,” *Journal of Political Economy*, Vol. 104, No. 5, pp. 958–978.
- Price, Joseph and Justin Wolfers (2010) “Racial discrimination among NBA referees,” *Quarterly Journal of Economics*, Vol. 125, No. 4, pp. 1859–1887.
- Thiele, Veikko (2013) “Subjective performance evaluations, collusion, and organizational design,” *Journal of Law, Economics, & Organization*, Vol. 29, No. 1, pp. 35–59.
- Tirole, Jean (1986) “Hierarchies and bureaucracies: On the role of collusion in organizations,” *Journal of Law Economics & Organization*, Vol. 2, p. 181-214.
- 北田智久(2015) 「主観的業績評価研究に関する現状と課題：文献レビューに基づく考察」, 『原価計算研究』, 第39巻, 第1号, 54-64頁.