

少数の人から消費行動で「参考になる」意見を引き出すには？

藤崎樹^a 本田秀仁^b 植田一博^c

要約

消費行動の際、私たちは他人の評価をよく参考にする。しかし、こうした他人の評価は時に自分の評価から大きくズレるものであり、消費の際の参考にならないことがある。すでに、評価者数が多いほど、評価の「参考になる度合い」が上昇することが示されているが、日常的には多数の評価者はまず手に入れられない。そこで本研究では、評価者が少数の場合を念頭に、より参考になる評価を引き出す手法を提案する。具体的には、刺激に対する評価者自身の評価に加え、他人の評価を想像し、あわせて回答させる。これにより、1人から擬似的に2人分の評価が得られると考えた。行動実験を通じて本手法を適用し、素材となる評価を収集した。これを基に、計算機シミュレーションを通じて評価の参考になる度合いを分析し、本手法の有効性を検証した。結果、評価者が少数の際、自身の評価のみの場合に比べ、本手法がより参考になる評価を引き出すことが明らかとなった。

JEL 分類番号： C63, C91, D83

キーワード： 口コミ, 情報探索, 集合知

^a 東京大学大学院総合文化研究科博士後期課程 bpmx3ngj@gmail.com

^b 安田女子大学心理学部講師 hitohonda.02@gmail.com

^c 東京大学大学院総合文化研究科教授 ueda@gregorio.c.u-tokyo.ac.jp

1. イントロダクション

「どの CD を買おうか」「どのホテルに泊まろうか」といった日常的な消費行動において、私たちは、商品のスペックやコストなどといった情報だけでなく、友人などでそれをすでに消費した人（以下、レビュワーと表記）の評価を参考にすることがある。こうしたレビュワーの評価は、自分（が消費した場合；以下消費者と表記）の評価とある程度近いことが多く、消費において「参考になる」ことが主張されている（Gilbert et al., 2009 など）。しかし一方で、レビュワーと消費者の評価（好み）は大きくズレてしまうケースもあり、こうした場合、レビュワーの評価を参考にすることは好ましくない消費をもたらしてしまう（「参考にならない」といえる）。

それでは、どのようにすればより参考になる評価を手に入れられるのだろうか。先行研究（Müller-Trede et al., 2017 など）では、レビュワーの人数が多いほど、評価の「参考になる度合い」が上昇することが定量的に示されている。研究は以下の形で行われた。行動実験を通じて、刺激（楽曲）についての参加者の評価を収集する。その上で、計算機上で参加者からレビュワーと消費者とをサンプリングする。そして、レビュワーの評価が消費者の評価とどれくらい近いか、すなわち、消費者の評価をどの程度予測できているか計算することで、評価の参考になる度合いを計測した。

とはいえ、こうした多数のレビュワーによる評価は、日常的にはまず手に入れることができない（Analytis et al., 2018）。

2. 目的

そこで本研究では、特にレビュワー数が少ない場合を念頭に、レビュワー自身の評価のみの場合に対し、消費者にとってより参考になる評価を引き出す手法を提案することを目的とした。具体的には、商品に対するレビュワー自身の評価（以下、‘Own’と表記）に加え、「他人」の評価を想像し（‘Guessed’）、あわせて回答させる。これにより、1人のレビュワーから擬似的に2人分の評価が得られ、より参考になるものが生まれると仮説を立てた。想像する「他人」としては、自身の考えとは異なる思考が促されることが広く知られている「世間一般」（Fujisaki et al., 2017 など）を設定した。

Own のみの場合に比べ、Own と Guessed とを組み合わせたもの（‘Double’、分析では2つの評価を平均化する）が消費者にとってより参考になるものとなるかどうか、計算機シミュレーションを通じて検討を行う。評価に関するデータは行動実験によって収集した。

3. 行動実験を通じた評価に関するデータの収集

3.1. 実験参加者

大学生・大学院生 56 名が実験に参加した。

3.2. 実験群

本手法を適用する 1 群のみ。

3.3. 実験刺激

先行研究 (Müller-Trede et al., 2017) と同様のものを用意した。具体的には、クラシックから民族音楽まで、幅広いジャンルの楽曲計 24 曲を、いずれも 1 分ほどの長さにカットしたものを各参加者に与えた。刺激の呈示順は、参加者ごとにランダム化した。

3.4. 実験手続き

実験は全て PC 上で行った。楽曲を聴かせた上で、どれくらい好きか 100 段階で評価させた (1 = 全く好きではない; 100 = 非常に好きである)。加えて、同一画面上に「世間一般の人はどれくらい好きだと思うか」を想像させる教示を呈示し、あわせて回答させた。この手続きを、24 曲全てに対して行った。

4. 計算機シミュレーションによる分析

先行研究 (Müller-Trede et al., 2017) と同様に、計算機上で選ばれた「レビュワー」が、「消費者」にとって参考になる度合いを分析した。参考になる度合いの指標としては、両者の評価のズレの大きさを表す MSE (Mean Squared Error; 評価値の差の二乗) を採用した。以下、まずレビュワーが 1 人の場合、本手法が、評価の参考になる度合いを向上させることを示す。続いて、レビュワーが複数人の場合、本手法がどこまで有効なものであるか検討を行う。

4.1. レビュワーが 1 人の場合

1 人のレビュワーと 1 人の消費者の組み合わせ、全 1,540 通り (= $56 * 55 / 2$) について、全 24 問の平均 MSE を計算した。レビュワーの評価としては、Own, Guessed, Double 全てが試された。図 1-A に結果を示した。分析の結果、Own と比べて、Double で評価のズレが小さいことが明らかとなった (95%信頼区間¹: Own = [812.74, 854.59], Double = [635.31, 668.74])。したがって、本手法を適用することで、1 人のレビュワーからより有益なレイティング情報を引き出すことができたと捉えられる。また、Guessed も Own よりも評価のズレが小さく (95%信頼区間: Guessed = [668.15, 705.21])、Double と Guessed に有意な差は見られなかった。この点は「5. 総合討論」で考察する。

¹ 95%信頼区間は、全てブートストラップ法を用いて求めた。具体的には、繰り返しを許すリサンプリングを 2,000 回行い、ここで得られたサンプルを算出した。

また、評価値の確率密度関数をあわせて表示した（図 1-B）. このグラフは、全参加者の全推定をまとめて表示している. 図より、Gussed, Double に比べ、Own で値の分散が大きいことが見て取れる.

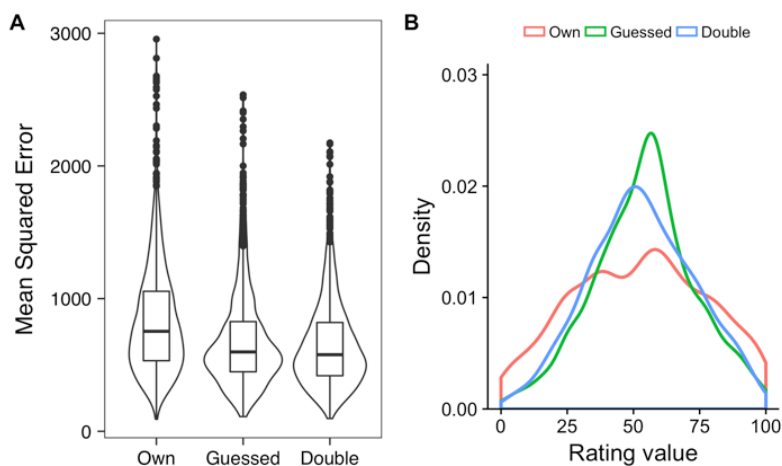


図 1. 分析 1 の結果. A : レビュー者の各推定における Mean Squared Error の比較. B : 全参加者の各推定の確率密度関数. Double は 2 つの推定を平均化したものを表示.

4.2. レビュー者が複数人の場合

先行研究 (Müller-Trede et al., 2017) に従い、計算機シミュレーションを通じて分析を行った (図 2). シミュレーションの手続きは以下の通りである.

【Step 1】 レビュー者数に応じて、全参加者からレビュー者がランダムにサンプリングされる (サンプリングは重複を許す).

【Step 2】 レビュー者の評価の平均値と、各参加者の評価値との MSE をそれぞれ計算する.

【Step 3】 Steps 1-2 をレビュー者数ごとに 2,000 回繰り返す.

【Step 4】 Steps 1-3 をレビュー者の評価 (全員 Own, 全員 Gussed, 全員 Double) ごとに行う.

図 3 にシミュレーションの結果を示した. レビュー者の人数としては、1 人から全 56 人までの全人数を分析対象としている. まず、Own と Double について、レビュー者の人数が増えるにつれ、評価のズレの差が縮まってしまうことが分かる. しかし、差がほぼ 0 となる 20 人以降、両者の値はほぼ同じであった.

それに対し、Gussed はレビュー者数の増加に伴う評価のズレの減少がゆるやかであった. 特にレビュー者数が 8 人以降、Own に対して値が大きくなってしまいうことが明らかに

なった.

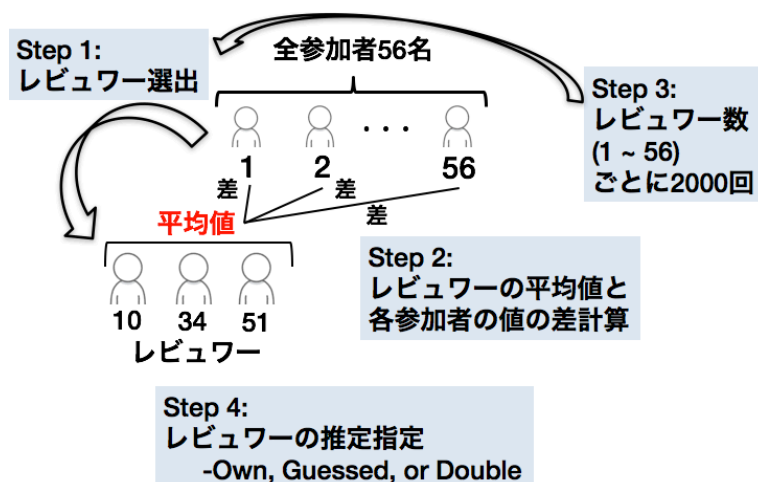


図 2. シミュレーションの手続き. イラストはレビュー者が 3 人の場合 (参加者番号 10, 34, 51) を例示している.

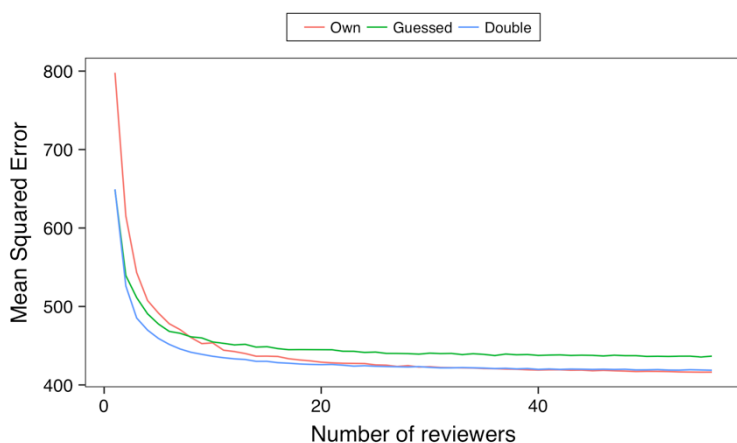


図 3. レビュー者が複数人の場合の結果.

5. 総合討論

消費行動の際, 他人の評価は有力な情報源といえる. とはいえ, 他人と自分の評価 (好み) はしばしば大きく異なる場合. 評価者の人数を増やすことで, 評価の「参考になる度合い」は上昇することが知られているが, 日常生活において多数の人の評価はまず集められない.

そこで本研究では, レビュー者が少数の場合を念頭に, レビュー者自身のみの評価に対し, より参考になる評価を引き出す手法を提案し, その有効性を行動実験と計算機シミュレーションを通じて検討した. 結果, レビュー者が少数 (具体的には, 20 人未満) の際に, 本手法がより参考になる評価を引き出すものであることが明らかとなった. よって, 日常的な消費行動の際, 本手法を適用することで相手からより参考になる評価を手に入れる (あるい

は与える) ことができると捉えられる。なお、レビュワーが1人の場合、想像した他人の評価のみと、本手法が提唱する2つの評価の組み合わせとで有意な差は見られなかったものの、レビュワー数が増えるにつれ、本手法の相対的な有効性が上昇した。したがって、本手法は参考になる評価を頑健に引き出すものであると捉えられる。

今後の課題としては、本手法がどのような場合にどこまで有効なのか、その条件を明確にすることが挙げられる。すでに、他人の評価の参考になる度合いは、刺激に対する familiarity (Müller-Trede et al., 2017) や、レビュワーと消費者の評価の similarity (Analytis et al., 2018) によって変化することが報告されている。本手法の有効性がこれらの要因によってどのように左右されるか、詳細に検討を行う予定である。

引用文献

- Analytis, P.P., Barkoczi, D. and Herzog, S.M, 2018. Social learning strategies for matters of taste. *Nature Human Behaviour* 2, 415-424.
- Fujisaki, I., Honda, H. and Ueda, K, 2017. On an effective and efficient method for exploiting "wisdom of crowds in one mind". *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, 2043-2048.
- Gilbert, D.T., Killingsworth, M.A., Eyre, R.N., and Wilson, T.D, 2009. The surprising power of neighborly advice. *Science* 323, 1617-1619.
- Müller-Trede, J., Choshen-Hillel, S., Barneron, M. and Yaniv, I, 2017. The wisdom of crowds in matters of taste. *Management Science* 64, 1779-1803.