

オンラインチャットのアカウント停止がアカウント停止者とその仲間の違反行為に及ぼす効果

横谷謙次[†] 高野 雅典[‡]

徳島大学創成科学研究科[†] 株式会社サイバーエージェント Media Data Tech Studio[‡]

1. 問題と目的

オンラインのチャット上では定期的に不適切な利用者（例えば差別的な言動を繰り返す利用者など）のアカウントを停止している。なぜなら、この種の不適切な利用は深刻な人権侵害につながる可能性があるからである（al-Khateeb et al. 2017; Kopecký, 2017）。しかし、アカウント停止が利用者のその後の違反行動に与える影響についての研究は、限られている（Chandrasekharan et al., 2020）。そこで、本研究では、アカウント停止の効果を検証するために、シミュレーションにより対照群を作成し（Bond et al., 2012; Yokotani and Takano, 2021b）、その対照群と実際の群とを比較することにより、アカウント停止の効果を検証する。

本研究の理論的枠組みは、罰の代理学習（Bandura et al., 1963）である。この理論では、仲間が違反行為を行って罰を受けている場面を目撃した人は、たとえ自分が直接罰せられていなくても、同じ違反行為を行う可能性が低くなると想定している。罰の代理学習は、オンラインネットワークにおいても確認されている。例えば、違法ダウンロードで処罰された場合、そのネット仲間は自分が直接処罰されていなくても、違法ダウンロードをやめることが知られている（Sun et al. 2013）。実際、差別表現を使っていた利用者がアカウント停止された場合、そのアカウント停止者の仲間も差別表現を減らすことが報告されている（Jhaver et al., 2021）。

これらの先行研究に基づき、アカウント停止を受けた者とその仲間は、シミュレーション上ランダムにアカウント停止を受けた者（ランダムな停止群）とその仲間よりも、違反行為のリスクが減少すると仮定した。また、違反者の人気は違反行為と正の相関がある可能性がある（Schwartz et al, 2006; Wegge et al, 2016）ため、違反者の人気をアカウント停止効果の媒介因子として推定した。

2. 方法

2.1. 対象者：調査協力の候補者は、2020年1月1日から6月30日までのピグパーティ利用者で、先行研究と同じ（Yokotani and Takano, 2021a）である。ピグパーティは日本のオンラインチャットアプリケーションであり、ユーザーは匿名の名前、アバター、個室を持つことができる（Takano and Tsunoda, 2019）。年齢・性別は不明であるが、過去の調査から10代～20代の女性がメインユーザーである可能性がある（Yokotani and Takano, 2021c）。候補者のうち、違反者または違反者の仲間である263,491人のピグパーティユーザーが本研究の対象者である。

2.2. 違反者とアカウント停止者：違反者は、ピグパーティで違反行為を行った利用者を示す。人間の評価者が、投稿の文脈を考慮し、同社が作成したガイドラインに従って、違反やペナルティ付与（それに伴うアカウント停止）を判定する。違反行為、24時間アカウント停止、48時間アカウント停止、永久アカウント停止は一日平均して、1,001回、52回、12回、5回起きている。また、違反者は28,305名であり、この内、24時間のアカウント停止者が8,357名、48時間のアカウント停止者が2,035名、永久アカウント停止者が899名である。

2.3. 違反者の仲間関係：ある利用者が違反者と1対1のやりとりをピグパーティで行った場合、その利用者は違反者と私的な関係を持っており、私的な仲間と推定した。同様にその利用者が集団、もしくは、オープンスペースでのやりとりをした場合、チーム的な及び公的な関係があり、仲間とされた。違反者の私的な関係、チーム的な関係、及び公的な関係はそれぞれ2,891,645、4,819,011、2,824,593である。

2.4. 違反行為の発生確率：まず、アカウントの停止前、停止中、停止後をそれぞれ、最初の停止開始前、最初の停止終了から最後の停止開始までの期間、最後の停止終了後の期間とそれぞれ定義する。次に、停止前の違反行為の確率を、停止前の違反行為数/全違反行為数として算出する。同様の手順で、アカウント停止中及

Effects of suspensions on offences of suspended offenders and their peers on an online chat platform. [†]Kenji Yokotani, Graduate School of Science and Technology for Innovation, Tokushima University. [‡]Masanori Takano, Media Data Tech Studio, CyberAgent, Inc.

び停止後の違反行為の確率を推定する。

2.5. 違反行為の潜伏期間: 違反行為の潜伏期間は、違反行為と違反行為の間、あるいは違反行為から打ち切り日(調査開始日、調査終了日、最初のアカウント停止日、最後のアカウント開始日)までの期間として推定した。少なくとも1回の違反を行った違反者は、アカウント停止前、停止中、停止後において潜伏期間が推定できる。

2.6. シミュレーションによるランダムなアカウント停止: このシミュレーションでは、ネットワークの構造、違反行為の頻度や期間は固定し、アカウント停止の期間と頻度のみをアカウント停止者にランダムに割り当てた。言い換えれば、アカウント停止者と停止時期や停止回数を紐づけた辞書を作成し、その辞書での紐づけをランダムに変更したが、その他の特徴は変更していない。24時間、48時間、永久停止者のランダムな辞書をそれぞれ200個作成した。この辞書とネットワークから、アカウント停止者とその仲間のアカウント停止後の違反確率と潜伏期間を推定した。これらの確率と潜伏期間は相互に独立であるので、正規分布が仮定出来る(Bond et al., 2012; Yokotani&Takano, 2021b)。そこで、それらの95%信頼区間を推定し、実際に観測されたアカウント停止者とその仲間の違反確率と潜伏期間が、ランダム辞書の95%信頼区間を超えているかどうかを確認した。

3. 結果と考察

3.1. 利用者の人気とアカウントの停止効果: アカウント停止者の人気は、アカウント停止が及ぼす当人の違反確率に対して負の媒介因子であった。つまり、人気の高いアカウント停止者は、アカウント停止後に、違反行為を行う確率が高かった。一方、彼・彼女らの人気は、アカウント停止による仲間の違反行為に対しては正の媒介因子であった。つまり、人気者のアカウント停止によって、その仲間は、その後の違反行為を行わなくなった。

人気者のユーザーは違反行動を行うことによって、人気を獲得している可能性があるため、アカウントの一時停止によって、更に違反行為の可能性が高くなることが推定された。違反行為の代理学習が容易に起こることを考えれば(Yokotani and Takano, 2021a)、こういったユーザーに対しては永久停止をすることが望ましいだろう。

3.2. アカウント停止が停止者の違反行為に及ぼす影響: 仮説通り、24時間および48時間のアカウント停止によって、その停止者の違反確率は、ランダムな停止群よりも有意に低くなった。

さらに、アカウント停止者がアカウント停止前に違反を犯す確率は、ランダムな停止群よりも有意に高かった。

また、仮説通り、アカウント停止者の停止後の違反行為の潜伏期間は、24時間、48時間を含めて、ランダムな停止群よりも有意に高くなった。同様に、アカウント停止者の停止前の違反までの潜伏期間は、24時間、48時間を含めて、ランダムな停止群よりも有意に低かった。これらの結果から、アカウント停止処分によって、停止者の停止後の違反確率を減らし、かつ、違反までの潜伏期間を伸ばせる、と言える。

3.3. アカウント停止が停止者の仲間の違反行為に及ぼす影響: 予想通り、24時間、48時間、永久停止のアカウント停止者の私的な仲間は、ランダムな停止群の仲間よりも停止後の違反確率が有意に低かった。同様に、24時間、48時間、永久停止のアカウント停止者の仲間は、ランダム停止群の仲間よりも停止後の潜伏期間が有意に高かった。また、同様の効果は、チーム的な仲間や公的な仲間においても、確認された。これらの結果は、すべてのアカウントの停止処分は、アカウント停止者の仲間の違反行為の確率を減少させ、違反行為の潜伏期間を増加させる効果があることを示している。

3.4. アカウント停止が、アカウント停止処分を一度も受けたことの無い仲間の違反行為に及ぼす影響: 代理学習の効果を明らかにするために、アカウント停止処分を全く受けたことのない仲間へのアカウント停止処分の効果を推定した。その結果、24時間、48時間、永久停止を含むすべてのアカウント停止処分は、停止処分歴の無い私的な仲間の停止後の違反行為の確率を低下させることがわかった。さらに、すべての停止処分は、彼・彼女らの停止後の違反行為までの潜伏期間を増加させた。また、同様の効果は、グループ的な仲間や公的な仲間においても、確認された。

4. 結論

本研究の結果から、アカウント停止は停止者だけでなく、その仲間の違反行為を減らす効果があると考えられた。

5. 謝辞

本研究の内容は Yokotani & Takano, Effects of suspensions on offences/damages of suspended offenders and their peers on an online chat platform. Telematics and informatics, in press と重複しています。また、本研究は株式会社サイバーエージェントより資金援助を受けています。