Fall 12-20-2016

How Should We Motivate Effort

Shamima Khan
CUNY Hunter College

How Should We Motivate Effort

by

Shamima Khan

Submitted in partial fulfillment
of the requirements for the degree of
Master of Arts Economics, Hunter College
The City University of New York

2016

Thesis Sponsor:

December 20, 2016	Karna Basu
Date	First Reader
December 20, 2016	Ingmar Nyman
Date	Second Reader

Table of Contents

I. Introduction

The relationship between incentive and effort is a major topic of inquiry in microeconomics. This relationship matters for the labor market outcome, whether we analyze it through the lenses of economic development, business management, or public policy. The carrot-stick reward system has been an integral part in the labor market, but it is not always the best motivator for optimum labor output. One of the big assumptions in economic models is that human beings are completely rational. Behavioral economics suggests otherwise – human beings are predictably irrational. This paper investigates how the relationship between incentive and effort is affected by human rationality and irrationality. Specifically, what incentive structure encourages people to give their highest effort? What incentive structure balances the thin line between frustration and aspiration?

This research builds up on goal-setting theory, which states that the closer someone is towards reaching the goal the more effort they tend to put in. This paper attempts to build up on that idea and investigates what happens when the goals are set in greater frequency and in different patterns. Does it significantly increase effort? If we vary the frequency of the incentives, does one motivate more than the others?

Be it marketing or the healthcare system, insight from this research could be used to revamp the incentive structure for consumers or service providers, like healthcare practitioners. It could make the system more efficient by saving thousands of dollars and time.

To simulate the real world assignments closely, all college students participating in the experiment were given choices of four pieces of either candies, pens, or pencils for completing 40 simple math questions. Keeping the test content and the total number of rewards the same, the five treatments were only varied by the patterns in which participants were given the rewards, with the fifth treatment used as a control with no rewards for the participants. Interestingly, the result from

this experiment went counter to the paper's main hypothesis that greater frequency would lead to more completion rate. Treatments with greater frequency of rewards had a lower completion rate than the treatment with all rewards at the end.

II. Lit Review

Several studies have been done to test how positioning of rewards affect performance or effort. In their paper on goal gradient hypothesis, Kivetz et al. (2006) did field experiments to test "the classic finding from behaviorism that animals expend more effort as they approach a reward" (Kivetz, Urminsky, & Zheng, 2006). They found that as participants get closer to earning free coffee or reward certificates after a certain number of purchases, their purchasing frequency accelerates. This acceleration is induced by the illusion of progress to the goal. Researchers then captured this relationship through a parsimonious function in which "effort investment is a function of the proportion of original distance remaining to the goal." This paper builds up on their findings to investigate what happens when the rewards are broken down into smaller pieces and the distances between them are varied through different treatments. According to their hypothesis, treatment with the smallest distance between the goals, should be most motivating, followed by the ones with bigger and then the biggest.

Debraj Ray in his paper "Aspirations, Poverty, and Economic Change" identifies "best aspirations" as the ones that "lie at the moderate distance from current situation" (Ray, 2003). The distance should be large enough to motivate them, but not too large to induce frustration. Based on Ray's findings, this paper hypothesizes that treatment with the largest distance will be least motivating as it will most likely induce frustration.

When it comes to incentives, the relationship between extrinsic and intrinsic motivation is an important one and has been widely studied. One paper studied this phenomenon indicating some data that suggests "that intrinsic motivation is diminished when extrinsic motivation is given—a

process known as the overjustification effect" (Greene, Sternberg, & Lepper, 1976). Another paper found that unless extrinsic motivation is properly done, overtime it can reduce intrinsic motivation for that behavior, because instead of working to get an incentive for the desired behavior, the external incentive undermines a previously held intrinsic motivation (Deci et al. 1999). This can then result in dependence on the extrinsic motivation for continued effort. However, if there is no such expectation, "and the extrinsic motivation is presented as a surprise, then intrinsic motivation for the task tends to persist" (Deci et al. 1999). This paper further tests the relationship between the motivation conflict with Kivetz et al. and Ray's findings by comparing other treatments with a treatment that has no rewards at all. Thus, the hypothesis that stems from this is that people with a knack for the test content will be inherently motivated to complete it, so incentives will make little to no difference in motivation, if not a negative one.

III. Experiment Design and Methodology

Since goal-setting theory mentions challenging and specific goals require increasing effort, this experiment is designed to test how much effort subjects put into taking a set of 40 relatively easy math questions, in increasing level of difficulty. There were five versions of an online form – each with a different incentive structure. Subjects were asked to pick a paper out of a hat which had the treatment letter hidden in it. Depending on which letter from A – E they picked, they took that form on an iPad provided to them. The incentives were given at the end of each pre-defined intervals to test what leads to the highest level of completion. Each treatment had the same 40 questions which were automatically graded and timed by the survey website. The total amount of incentive at completion were the same (four items of their choice) regardless of the treatment they picked, excepting treatment E which provided no rewards.

The five treatments are: Participants taking treatment A received all the rewards at the end of all 40 questions, and anyone stopping before completing all 40 were not given any rewards.

Participants of treatment B were given incentives at two equal intervals: they had to complete the first 20 questions to get the first two rewards and the last 20 questions to get the other two rewards. Participants taking treatment C received incentives in four equal intervals, i.e. upon completing every 10 questions they received one reward. In treatment D, people received rewards at random intervals. The random interval was chosen by the participants by picking another folded paper with the interval written inside the fold, but they did not know it until they had reached that endpoint. The closer their endpoint to the 40^{th} question, the higher the reward. For example, if someone picked "30", the participant was given three items of their choice, whereas someone who picked "10" was only given one item and none afterwards if they continued to do more. Finally, treatment E was a control where participants taking it did not receive any rewards no matter how much of the test they completed.

The treatment forms were given to anonymous Hunter college students. The process for participants was as follows: 1) Approach Hunter College students in its main campus. 2) After they were presented with the pitch, if they were willing to participate, they were given a consent form to sign. 3) Each of the willing participants were asked to choose one slip from five of the folded paper slips (each containing handwritten link address to different forms). For example, if a participant picked form B, they could then access the form through their device or the iPad provided them to them for the task. 4) Students were then rewarded according to the incentive structure of the form they picked. Participants only received rewards only when they completed all the questions up to that endpoint.

Subjects were free to quit the survey anytime. The incentives were choices between different types of candies, and multicolored pens and pencils. The study was administered around

the Hunter main campus, particularly outside the library and in the lounge areas. Participants were graded automatically as they answered the electronic quiz.

Variables from the Experiment:

	Name	**Stata Code**	**Coding description**
1	Treatment	treatment	A = all rewards at the end B = rewards at 50% interval C = rewards at 25% interval D = rewards at random interval F = control
2	Participant id	Id	Randomly assigned
3	Time spent in minutes	minspent	Time spent taking survey
4	Gender	male	Male = 1 Female = 0
5	Major	major	Analytical (math, economics, finance) Lifesciences Social Sciences Arts Other (including undecided)
6	Education	education	Freshman = 1 Sophomore = 2 Junior = 3 Senior = 4

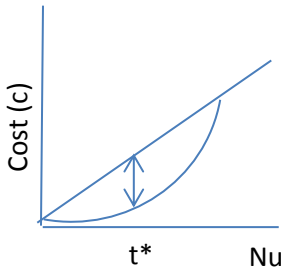
			Graduate = 5 Other = 6
7	Time of the day survey taken	responsetime	Time: Morning = 1 (until 12 pm) Noon = 2 (12 pm – 3 pm) Afternoon = 3 (3 pm – 6 pm) Evening = 4 (6 pm – 10 pm)
8	Total Q answered by each responder	totalans	Continuous (1-40)
9	Met need for treatment D?	fdmetneed	Lottery Choices to gain rewards: 25% 50% 75% 100%
10	Completion – 100%	hundred	Discrete (0= no, 1=yes)
11	Completion – 75%	seventyfive	Discrete (0= no, 1=yes)
12	Completion – 50%	fifty	Discrete (0= no, 1=yes)
13	Completion – 25%	twentyfive	Discrete (0= no, 1=yes)
14	Completion – >25%	none	Discrete (0= no, 1=yes)
15	Completion rate	comprate	= (# question answered/40)*100
16	Completed?	probcomp	Yes=1, No=0

IV. Theoretical Predictions

This paper investigates three hypotheses. First, basing on the findings that people put in more effort as they get closer to a goal, smaller intervals of rewards incentivize people to get to the goal faster. Second, targeted well-defined incentives are more effective than random incentives. Third, people with an intrinsic motivation towards the task at hand will be indifferent or act negatively towards the reward being given.

According to the standard model in economics, it is assumed that human beings are perfectly rational (utility maximizer), have perfect information (external knowledge), and know their preferences well (internal knowledge). As a result, as the number of tasks increases, it is expected that their cost of performing the task increases. Therefore, a rational agent will continue to do the task as long as their cost of doing the task equals the benefit from the reward, at an optimum number of task, t*.

Maximizing rewards:



$$r(t) = \propto t$$

$$\max \propto t - c(t) \Rightarrow \text{reward} - \text{cost}$$

$$\text{FOC} \propto = c'(t)$$

Hypothesis: $\propto$ goes up, t* goes up

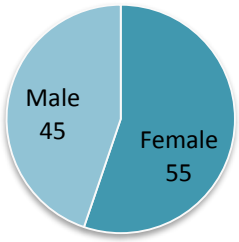
According to the behavioral model in economics, if the participants are risk averse, they will have higher completion rate with smaller intervals (treatment B and C) and lowest completion rate with treatment D since the reward for that is unpredictable.

V. Results

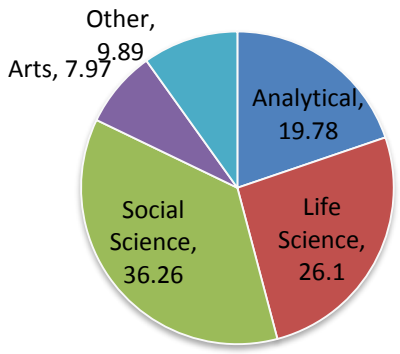
Demographics

There was a total of 350 observations in this pilot study, with 5% more female compared to male. Per treatment there were 70 participants. Their concentrations of studies were bucketed into five categories: social sciences, life sciences, analytical, arts, other. Since the survey is based on math questions, a separate category called analytical is used. Numbers-heavy concentrations such as math, economics, finance, business, accounting etc. are categories under this bucket leaving behind other concentrations such as education, religion, English etc under the category of social sciences. The highest number of participants were in the social sciences, 36%, followed by life sciences, 26%. Participants' levels of education were bucketed into six categories, with highest number of participants being students who graduated or are in an advanced degree after their bachelors.

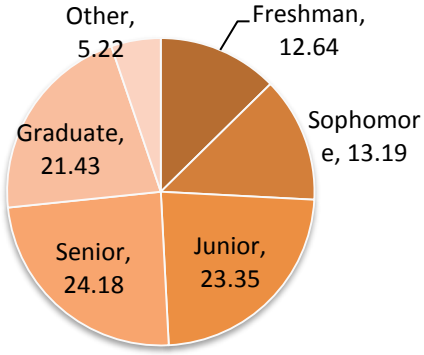
1. Percentage of Participants across Gender



2. Percentage of Participants across Majors

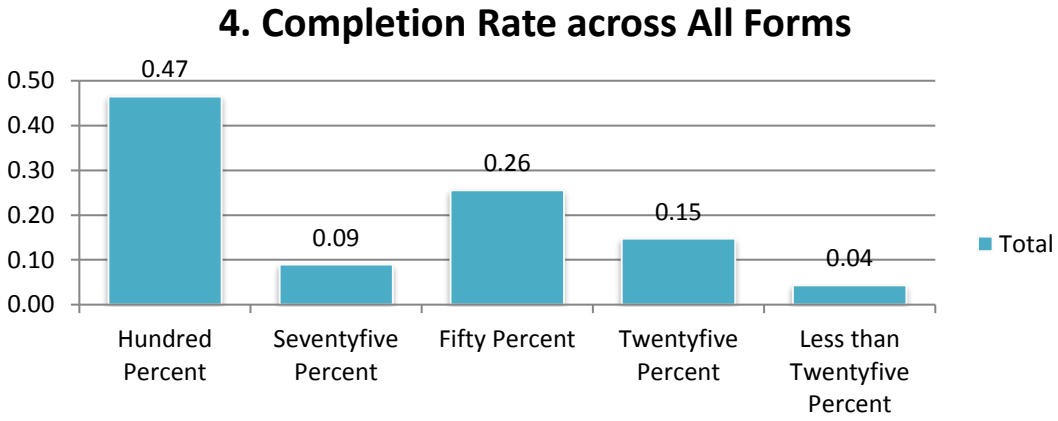


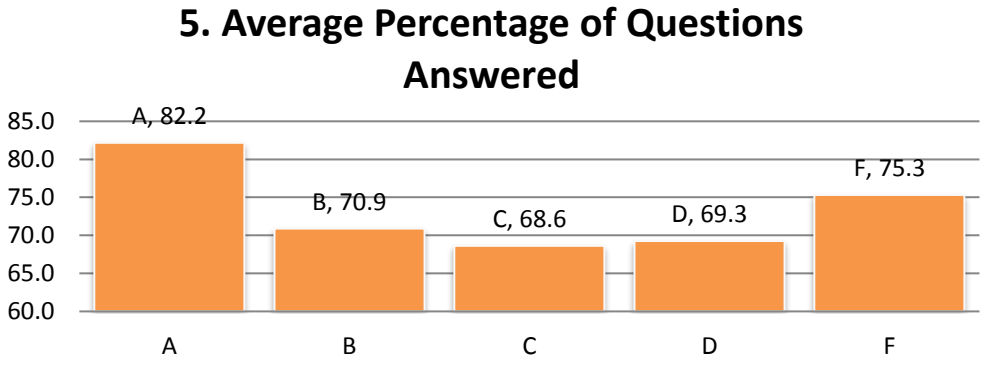
3. Percentage of Participants across Education Levels



Outcomes

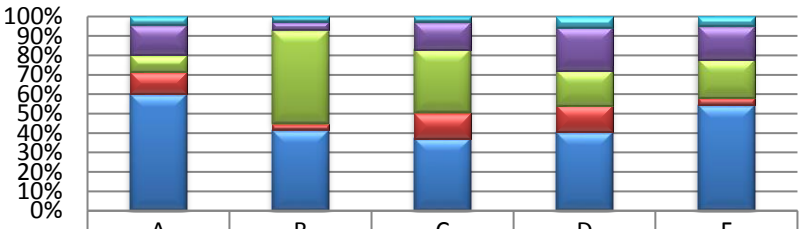
Overall, 47% of the participants completed all 40 questions in the study, followed by 26% who completed half of the questions. There were some clear nudges that show up in the aggregated completion rate. In figure 4, across treatments, in terms of the highest average percentage of questions answered, treatment A performed best (82.2%) where all rewards were at the end, followed by treatment F (75.3%) where there was no reward. Treatment B and D performed roughly equally. Treatment C had the poorest result (68.6%) which had the highest breakdown of rewards, this is outcome is slightly lower than treatment D.





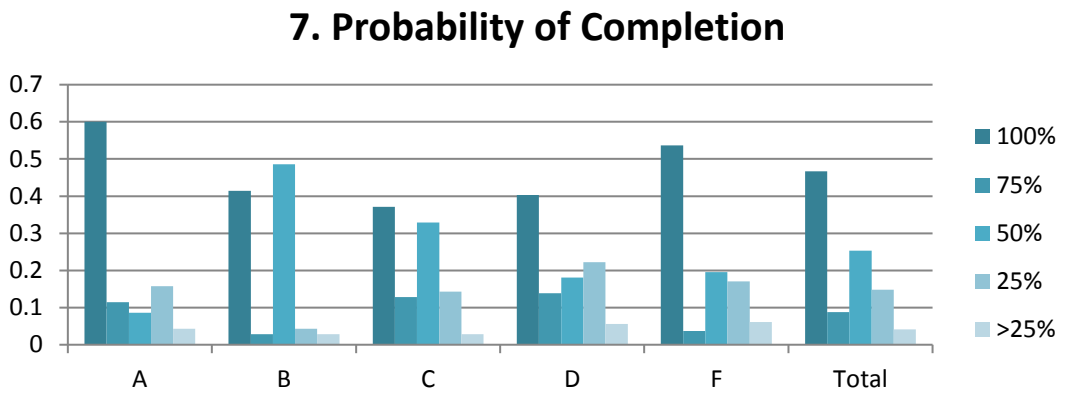
According to figure 6, the percentage of people who gave up before completing one quarter of the task is small across all treatments. In treatment A where rewards were given all at the end, 60% of the participants completed all of the questions. In treatment B, where rewards were given in two intervals, the highest completion rates at the breakdown. Since treatment C had rewards in 4 intervals, the completion rates were relatively more widely broken down. The lottery system in treatment D shows most random breakdown, in terms of frequency. In absence of rewards, treatment F shows similar result to treatment A.

6. Completion Rate Across Each Form



	A	B	C	D	F
Less than Twentyfive Percent	0.04	0.03	0.03	0.06	0.05
Twentyfive Percent	0.16	0.04	0.14	0.22	0.17
Fifty Percent	0.09	0.49	0.33	0.18	0.20
Seventyfive Percent	0.11	0.03	0.13	0.14	0.04
Hundred Percent	0.60	0.41	0.37	0.40	0.54

As figure 7 shows, the overall probability of 100% completion for all participants is less than 0.5. The probability of 100% completion is highest in treatment A, followed by treatment F. As we can see the breakdown of the rewards clearly nudged participants ending their tasks at those intervals. In both treatment A and F, participants were more likely to complete the task once they started since, compared to other treatments, the probabilities are small at other quarters of the treatment. Treatment D has the most randomly distributed probability of completion.



VI. Regressions

	Completion rate	Completion rate	Minutes spent	Probability of completion
Treatment				
B	-12.31*	-11.65*	-1.69	-0.2*
C	-14.45*	-14.34*	-3.42*	-0.26*
D	-14.22*	-16.31*	-2.19	-0.22*
E	-7.49	-8.37	-3.26	-0.07
Response Time				
Noon	-	-10.54	-	-
Afternoon	-	-12.88	-	-
Evening	-	-2.58	-	-
Major/Concentration				
Arts	-13.31	-13.62*	0.07	-0.22
Life Sciences	-4.2	-4.24	-0.03	-0.1*
Other	-6.65	-5.81	-1.25	-0.13*
Other Social Sciences	-10.05*	-9.51*	-0.85	-0.17*
Education Level				
Sophomore	-0.28	-0.39	-0.14	-0.03
Junior	1.07	0.86	0.29	-0.02
Senior	4.96	3.9	2.28	-0.11
Graduate	-0.17	-0.36	-0.32	-0.05
Other	-1.21	-1.25	2.57	-0.03
Gender				
Male	6.67*	6.69*	-0.64	0.09

Completion rate: Controlling for participant's major, education, and sex, compared to treatment A, all the other treatments were worse at motivating effort. At 5% significance level, people with treatment C (incentives at four intervals) were least likely to complete the task. Although statistically insignificant, people from other majors were less likely to complete the task compared to people from analytical majors, compared to females, males had a higher completion rate.

Controlling for time of the day, the completion rate decreased in a similar fashion, i.e. completion rate decreased for other treatments when compared with treatment A. Compared to morning time, the rate of completion was lower at other times of the day, although this result is not statistically significant.

Time taken: Controlling for other factors, participants spent least time during treatment A where the rewards were all at the end and most time at treatment C and F. For C a possible factor could be the time taken to get the reward at increments which might have slowed them down.

More likely or less likely to complete: Controlling for other factors, at a statistically significantly number of participants from all other treatments are less likely to complete the task than participants of treatment A.

VII. Analysis

This experiment's findings prove the following for each of hypotheses under study: First of all, smaller intervals of rewards incentivize people to get to the goal faster -- under the hypothesis that people put in more effort as they get closer to a goal. This paper rejects the hypothesis, since participants were less likely to complete a task or at a faster rate when rewards were given in more frequent intervals.

Secondly, targeted well-defined incentives are more effective than random incentives. During treatment A, rewards at the end kept participants going on longer towards completion, while shorter incentives reminded them that they can end at any time since they already received some reward. This paper cannot completely reject the hypothesis. It is true that participants ended the task at more irregular intervals during treatment D than during any other treatments. Interestingly, they were less likely to complete than treatment F, with no rewards.

Thirdly, people with an intrinsic motivation towards the task at hand will be indifferent or act negatively towards the reward being given. This paper fails to reject the hypothesis, since treatment F, with no rewards, showed the second most completion rate. This goes against the notion of standard economic model which states agents are perfectly rational beings. Some possible explanations are: they were intrinsically motivated to do that task (in this case math questions) or they were altruistic in their effort to help a fellow student with her research. When rewards were introduced, the task became a means to an end (to get the rewards) rather than an end in itself (doing it because they enjoy it).

These findings have important labor market implications. Before getting to that discussion here are the three big limitations of this study:

First, this study has been carefully designed to mimic the real world labor market as much as possible. However, there were limitation as to how realistic was the study. For example, even though participants were offered a variety of items, both in terms of function and preference, to induce the open choices available with money, a lot of participants mentioned they would approach the test with a different mindset if money was the reward.

Second, given the number of treatments studied, the number of observations is not only small but also biased. Despite controlling for very many factors, such as education level and major, these students were surveyed at one college campus.

Third, unlike the real world where people have to work to earn rewards (money), participants in this study were also biased in the sense that people might have declined to participate because they experienced self-handicapping (Schwinger, Wirthwein, Lemmer, & Steinmayr, 2014). Schwinger et al. studied this phenomenon and found that some people avoided effort in order to prevent future failure from hurting one's self-esteem.

Extrapolating from these research findings thus bring to mind the following implications for the labor market: Firstly, for analytical tasks such as these math questions, it is better to give the rewards all at the end up as a lump-sum than in increments. A possible explanation for this could Steinkuhler et al.'s finding called escalation of commitment that states humans often increase their commitment to a prior decision hoping that this additional investment will lead to a successful outcome (Steinkuhler, Mahlendorf, & Brettel, 2014). Another possible explanation could be commitment bias as studied by Gopinath in which he finds some personalities have difficulty retreating from positions to which they publicly committed (Gopinath, 2009).

Another incentive design implication of the previous point is that if the desired result is to encourage people to do the test but it is not essential to complete it, then rewards at shorter intervals is a good idea.

This research proves that crowding effect is happening even in this smaller scale, as shown in the high completion rate of treatment F with no rewards. Therefore, if employers want to

encourage participants to bring a task to completion, it is important for them to design the task in a way that speaks to their intrinsic motivation, even if it means not mentioning the reward at all.

Finally, since treatment D offered rewards at random unknown intervals, this research proves that participants value certainty. Therefore, if employers want participants to complete a task it is important for them to be clear about the expectations from the task and rewards being offered. This finding is in conflict with Shen et al.'s findings called Motivating Uncertainty Effect. (Shen, Fishbach, & K. Hsee, 2015).

VIII. Conclusion and Further Research

To get statistically significant results that further prove the validity of the hypotheses presented in this paper, it is important to have a larger sample with better incentives.

During the experiment one key observation was the influence of the group setting in participants' performance. Participants varied in their performance based on the people they were sitting with before I approached them. The group dynamic played an important role in when and for how long they took the test. Observing the influence of this factor was beyond the scope of this experiment. It would be interesting to see whether the group dynamic showed a pattern of completion across all treatments.

Another interesting area of enquiry would to test the impact of different incentive structure, these treatments, on creative vs analytical tasks. How would the treatments influence motivation level if participants were assigned two these types of assignments?

Another variable not studied in the paper is the correlation between choice of rewards and completion rate. Since there were two categories of rewards: functional (pens and pencils) and consumable (candies), do participants' preference for either correlate with the type of performer

they are. For example, are candy-choosers more into instant gratification? If so, are they more or less likely to complete the task differently across all treatments?

IX. References

Gopinath, M. (2009). The effect of public commitment on resistance to persuasion: The influence of attitude certainty, issue importance, susceptibility to normative influence, preference for consistency and source proximity. *International Journal of Research in Marketing*, 60-68.

Greene, D., Sternberg, B., & Lepper, M. (1976). Overjustification in a token economy. *Journal of Personality and Social Psychology*, 1219-1234.

Heath, C. (1999). On the Social Psychology of Agency Relationships: Lay Theories of Motivation Overemphasize Extrinsic Incentives. *Organizational Behavior and Human Decision Processes*, 25-62.

Kivetz, R., Urminsky, O., & Zheng, Y. (2006). The Goal-Gradient Hypothesis Resurrected: Purchase Acceleration, Illusionary Goal Progress, and Customer Retention. *Journal or Marketing Research*, 39-58.

Ray, D. (2003). Aspirations, Poverty and Economic Change. *New York University and Instituto de An´alisis Econ´omico*, 1-12.

Schwinger, M., Wirthwein, L., Lemmer, G., & Steinmayr, R. (2014). Academic Self-Handicapping and Achievement: A Meta-Analysis. *Journal of Educational Psychology*, 744-761.

Shen, L., Fishbach, A., & K. Hsee, C. (2015). The Motivating-Uncertainty Effect: Uncertainty Increases. *Journal of Consumer Research Resource Investment in the Process of Reward Pursuit*.

Steinkuhler, D., Mahlendorf, M., & Brettel, M. (2014). How Self-Justification Indirectly Drives Escalation of Commitment -- A Motivational Perspective. *Schmalenbach Business Review*, 191-222.