Too Much Nudging: Can it Cause a Decrease in the Desired Response?

Jay Shah¹, Rohen Shah#, and Philip Liang#

¹Aditya Birla World Academy, Mumbai, Maharashtra, India.
#Advisor

ABSTRACT

Nudging is an important economic tool for governments as well as private corporations. While it is a common assumption that nudging can help maximize social benefit by altering choice architecture there is still much to research regarding the effects of nudging too much. In this paper, several mathematical models based on different assumptions have also been created to explain the hypothesis that excessive nudging can cause a decrease in the desired response. Psychological phenomena have been explored to posit reasons for this behavioural irrationality. After reviewing previous literature of experiments with multiple ‘doses’ of nudging, it is concluded that nudging too much may in fact have an adverse effect. A comparison to a Laffer curve derivation in this paper is the graphical model that future research may experimentally find to prove the hypothesis.

1 Introduction

According to Thaler and Sunstein (2008) a nudge is defined as "Any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates. Putting the fruit at eye level counts as a nudge. Banning junk food does not." In a rational world, nudging should not have any effect on choice as it is based the premise of human behavioural irrationality. Thus, it is plausible that too much nudging can cause a decrease in the desired response because of the aforementioned irrationality. For example, a person may be less responsive to a barrage of highly frequent text message reminders because of how they perceive it. There are many different types of nudges such as defaults, framing, social proof heuristics, etc. In this paper, we will be focusing on nudges that can measurably vary in quantity and frequency. Hence, we will primarily be focusing on repeated nudges such as prompts and reminders.

In this paper, I will look at the impact of repeated nudging on the desired response.

2 Motivation
Today, nudge theory is a prominent tool in shaping both public and private sector policies. Ever since the 2008 book Nudge: Improving decisions about Health, Wealth and Happiness, by Nobel laureate Richard Thaler and legal scholar Cass Sunstein, the research and popularity into nudge theory has grown exponentially (Thaler and Sunstein, 2008). On the national level, many countries have created “nudge units” as depicted in Figure 1 (OECD research, 2018) to further their domestic public policy goals while the cost-effectiveness of nudges has attracted increasing implementation from the private sector in both marketing and management (Beggas, 2016; Güntner et al., 2019). This increasing popularity is illustrated by the increase in Google searches of the term “nudging” worldwide shown in Figure 2.

**Figure 1:** Map of Nudge Units around the world

---

1 Image taken from https://pbs.twimg.com/media/DWpmOTAXUAE4JP2.jpg:large
2 Image taken from https://trends.google.com/trends/explore?date=all&q=nudging
Thus, any research on nudging can impact millions of people across the globe. Two primary reasons to be concerned with repeated nudging are cost and ethics.

2.1 Cost of Nudging
The definition of nudging implies a very low cost of implementation due to the fact that economic incentives for those nudged remain unchanged in nudging. This means that a financial benefit (a discount) on buying fruits as compared to junk food cannot be counted as a nudge, as the intervention changes the economic benefits for the consumer to buy fruits. The favourable administrative costs in comparison with taxes and other hard regulation make nudging a more cost-effective approach (Benartzi et al., 2017). However, nudging does have a cost and the more you nudge, the greater the cost. Hence, if too much nudging caused a decrease in the desired response, it would also cause an inefficient allocation of resources. This could be especially important for government run nudge units that have a budget.

2.2 Ethics of Nudging
Sunstein (2015) states that an ethical nudge must promote three ideals of welfare, autonomy and dignity of the individual and undermine none of them. If too much nudging leads to a decrease in the desired response or even a negative response, it could raise ethical concerns. Such concerns would undermine the welfare aspect of an ethical nudge as the nudge would decrease the welfare of the individual if it promotes the undesired response. An ethical nudge should encompass a choice architecture where those nudged are nudged towards a choice that increases their welfare (as judged by those nudged). If repeated nudges lead to a decrease or reversal in the response they were designed for then the welfare is actually decreased, voiding the first parameter that defines an ethical nudge. An example of this would be if a person was given flyers everyday at the gym that asked them not eat junk food with...
tempting pictures of the foods, it may lead to a subconscious desire to consume the food and that may affect the person’s choice when presented with one. A nudge that causes a decrease in the desire response may also undermine the dignity ideal if the undesired response leads to a loss of dignity for the individual.

Hence, the effect of a too much nudge is highly important on financial and ethical grounds as nudging has a widespread impact today.

3 Plausibility of the theory

The basis of nudge theory is on the belief that human actions, choices and preferences are inconsistent and influenced by various cognitive heuristics, biases, and mechanisms that would be irrelevant to a rational point of view. Thus, the irrationality in our decision-making is what led to the introduction and popular influence of nudge theory in the 21st century. However, this very irrationality may provide some answers to the research question. The phenomenon of psychological reactance and desensitization bring forth plausible reasons for a decrease in the desired response caused by too much nudging.

3.1 Psychological reactance

Reactance is defined as a cognitive or behavioral response occurring when an individual feels her freedom is threatened, e.g., by an external intervention (Brehm, 1966). Nudge theory and nudges have had various criticisms on limiting behavioural autonomy (Hausman and Welch, 2010; Rebonato, 2014; Sunstein, 2017). When these same criticisms were voiced by those nudged (Felsen et al., 2013) it lead to a reduction in the nudge effects and even create the opposite response. Psychological reactance can explain such reactions (Haggag and Paci, 2014; Costa and Kahn, 2013; Arad and Rubinstein, 2018; Hedlin and Sunstein, 2016). Across a series of four studies, researchers found out that unsolicited advice that contradicts the initial impression of the decision maker can lead to the activation of a reactance state where the decision maker moves to contradict the advice given (Fitzsimons and Lehmann, 2004). Repeated nudging may cause those nudged to be more aware of the nudges. As the nudged are consciously aware of the overt nudges they may act in the opposite direction of the nudge as their freedom is threatened. Hence, repeated nudging may cause the onset of psychological reactance to play a role in the effectiveness of the nudge.

3.2 Desensitization

Popularized by Mary Jones Cover in her famous study (Jones, 1924), desensitization is an important concept in various psychological processes and literature. Desensitization occurs when the response to a certain stimuli decreases over repeated exposure to the said stimuli over time. Various studies have shown the effect of desensitization on the response to a stimuli. For example, children that are exposed to TV violence are less aroused autonomically to violence in real life [?]. A nudge is equivalent to the stimulus and its repetition can be comparable to the desensitisation process. Hence, in the context of the research question the concept of desensitization provides a plausible causation for the decrease in effectiveness of each nudge following the first.

4 Comparison to mainstream economic theories

A decreasing response to repeated nudging is a counter-intuitive concept as the general assumption would be that the more one is nudged the more effective the nudges are. However, this idea has relevant comparisons to economic concepts such as the Laffer Curve (Laffer, 2004) and the theory of diminishing marginal utility (Mankiw, 2014).

4.1 Comparison with Laffer Curve
The Laffer Curve (as shown in figure 3\(^3\)) is a graphical display of the relationship between the tax rate a government imposes and the total tax revenue it collects. Formalized by Arthur Laffer, the theory provides basis for the argument that cutting taxes may actually lead to an increase in the tax revenue if the tax rate is beyond the optimum T* value.

Thus, similar to the Laffer curve in shape there may exist a curve that illustrates the relationship between the frequency of nudging and the effectiveness of the repeated nudge. This proposed curve is depicted in Figure 4. The curve initially has a decreasing positive gradient that reaches an optimum point for nudge effectiveness maximization (N*) and then follows an increasingly negative gradient. While the Laffer curve ends at 100% tax rate, the nudge effectiveness curve would continue below the 0 and into a negative effect if the nudge frequency was way too high.

However, the exact shape of the Laffer Curve is not known. That is because we are still unaware if any skews exist, positive or negative, in the curve’s shape or if there may exist a plateau section. Hence, the exact shape of the proposed nudge curve can only be created through further experimentation that can pinpoint the nudge effectiveness at various points of frequency.

The continually decreasing gradient of the Proposed Nudge Curve also supports the hypothesis that the first nudge is the most effective with subsequent nudges being lower and lower in their effectiveness.

\[\text{Image taken from https://www.investopedia.com/terms/l/laffercurve.asp}\]
4.2 Derivation of Laffer Curve

For the derivation of a simple Laffer curve in a single direct tax economy, we assume that the taxpayer has a utility function: \( \sqrt{c} + \sqrt{l} \) where \( c \) represents consumption and \( l \) represents leisure. The budget constraint is \( c = w(1 - t)(24 - l) \) where \( t \) represents the tax rate and \( w \) represents the wage rate. After substitution, the utility function is \( \sqrt{w(1 - t)(24 - l)} + \sqrt{l} \). The consumer maximizes utility by choosing \( l \) (the amount of time they spend on leisure i.e. not work) while the parameters \( w \) and \( t \) are taken as given constants. The first-order condition (the partial derivative of utility with respect to leisure) is

\[
0 = \frac{\frac{1}{2}}{w(1 - t)(24 - l)} \left( -w(1 - t) \right)^{\frac{1}{2}} + \frac{1}{2} \frac{1}{l^{\frac{1}{2}}}
\]

Thus, \( w(1 - t)(24 - l) \left( -w(1 - t) \right)^{\frac{1}{2}} = \frac{1}{l^{\frac{1}{2}}} \).

By squaring both sides and then taking the multiplicative inverse of both sides, we conclude

\[
\left( w(1 - t)(24 - l) \left( -w(1 - t) \right)^{\frac{1}{2}} \right)^{-2} = \left( \frac{1}{l^{\frac{1}{2}}} \right)^{-2}
\]

\[
= \frac{w^2(1-t)^2}{24-l} w(1-t)(24-l) = l.
\]

Thus, \( \frac{24-l}{w(1-t)} = l \). Therefore,

\[
\frac{24}{w(1-t) + 1} = \frac{1}{l} \left( 1 + \frac{w(1-t)}{w(1-t)} \right)
\]

\[
= \frac{w(1-t) + 1}{w(1-t)}.
\]

Thus, \( l = \frac{24}{w(1-t) + 1} \).

Therefore, revenue is \( t \left( 24 - \frac{24}{w(1-t) + 1} \right) \). If we assume that \( w = 1 \), then

\[
l = \frac{24}{24 - t + 1} = \frac{24}{24 - t}.
\]

Therefore, revenue is \( t \left( 24 - \frac{24}{24 - t} \right) \).

The following graph (Figure 5) is modeled from the derivation.
The same derivation can be applied to nudging. For example, if advertising is viewed as a nudge then consider a website that is trying to nudge people into buying products from its sponsors. \( l \) then is interpreted as time not spent on the website. \( t \) can be interpreted as the proportion of the time a user spends on the website that must be spent viewing advertisements. If advertisers pay the website \( p \) per unit of time that visitors spend viewing ads, the website’s revenue from the users is \((1 - l)t p\).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Laffer curve interpretation</th>
<th>Advertising interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( c )</td>
<td>Consumption</td>
<td>Valuable content viewed on the website</td>
</tr>
<tr>
<td>( l )</td>
<td>Leisure</td>
<td>Time spent not on the website</td>
</tr>
<tr>
<td>( t )</td>
<td>Tax rate</td>
<td>Proportion of browsing time occupied by advertising</td>
</tr>
<tr>
<td>( w )</td>
<td>Wage</td>
<td>Value of a unit of time spent viewing content</td>
</tr>
<tr>
<td>Revenue</td>
<td>( t \left( 24 - \frac{24}{w(1-t) + 1} \right) )</td>
<td>( t p \left( 24 - \frac{24}{w(1-t) + 1} \right) )</td>
</tr>
</tbody>
</table>

4.3 Comparison to the theory of diminishing marginal utility

The theory of diminishing marginal utility (as illustrated in Figure 5) states that the marginal utility derived from the consumption of an additional unit continually decreases, ceteris paribus. Hence, the consumption of the first unit gives the maximum marginal utility.

In comparison to this theory, the hypothesis proposed by this paper would suggest that an additional nudge in a given time period would have a diminishing effect and each successive nudge would have a lower marginal effect.

---

on the nudged. This would provide causation for the shape of the aforementioned proposed nudge effectiveness curve. The reason for this decreasing marginal effect could be psychological factors considered above or some other factor not yet considered.

**Figure 6:** Theory of Diminishing Marginal Utility

4.4 Quantified Marginal Effectiveness of a Nudge

Under the assumption that marginal effectiveness can be measured and quantified, the change in effectiveness of nudging of moving from $S_1$ to state $S_2$ is

$$\Delta E = E(S_2) - E(S_1)$$

The variable $n$ represents dosage (quantity or frequency) of nudges that is itself quantified. Under the assumption that $S_1$ and $S_2$ are only affected by the variable $n$ and holding all other variables constant, the marginal effectiveness of nudging is

$$\frac{\Delta E}{\Delta n}$$

Assume

$$\lim_{\Delta n \to 0} \frac{\Delta E}{\Delta n}$$

exists. Then, the marginal effectiveness of nudging is the first order partial derivative

$$\frac{\partial E}{\partial n} = \lim_{\Delta n \to 0} \frac{\Delta E}{\Delta n}$$

Thus, if $\lim_{\Delta n \to 0} \frac{\partial E}{\partial n}$ exists and $\frac{\partial^2 E}{\partial n^2} < 0$, then nudging has diminishing marginal effectiveness.

5 Alternative models

In this section, I have explored some other mathematical models that provide explanations for the decreased effect of increased nudging based on various assumptions.
5.1 Simple Backfiring Risk Model

With this model we assume that a nudge can cause a reversal in the desired behaviour to part of the population engaging in the desired behaviour. Thus the optimal amount/frequency of nudging would be the one that causes a maximum net benefit to society.

The parameters in this model are as described below.

- \( p \): proportion of population engaging in a desired behavior
- \( e \): effectiveness of nudge - proportion of people engaging in an undesired behavior that an effective nudge causes to engage in a desired behavior
- \( h \): probability that a nudge backfires i.e. causes a person engaging in a desired behavior to engage in an undesired behavior
- \( b \): benefit to society of a person engaging in a desired behavior rather than an undesirable behavior
- \( c \): total cost of creating a nudge

The nudge is effective with probability \( e \). This means that \( e \) is the proportion of people engaging in an undesirable behavior that the nudge persuades to engage in the desired behavior. The proportion of the population engaging in an undesired behavior is \( 1 - p \). So \( e(1 - p) \) is the proportion of the population that is convinced to engage in a desired behavior due to the nudge. The nudge backfires with probability \( h \). This means that \( h \) proportion of the population will stop engaging in the desired behaviour after being nudged. \( p \) is the proportion of the population engaging in the desired behavior. A nudge backfires and causes \( ph \) people to engage in an undesired behavior. The net benefit of a nudge is therefore:

\[
b(e(1 - p) - ph) - c.
\]

However, there must be ethical considerations if a nudge backfires as it would be unethical for society to value the net overall benefit if the benefit to an individual is negative. Thus, a nudge that can backfire i.e. cause an individual to reverse from the desired behaviour is almost always not implemented.

5.2 Information Utility model

This model assumes that nudges provide the agent with information that causes changes in their perceived utility. For examples \( u(c, s; r) \) is a utility function over calls while driving and safety. The choice variables are \( c \) and \( s \). \( c \) represents calls while driving and \( s \) represents safety. \( r \) is a parameter that represents the riskiness of taking a call while driving. The optimal level of \( c \) is decreasing in \( \mathbb{E}(r) \). The agent initially has some belief about \( \mathbb{E}(r) \). Providing information that convinces drivers that distracted driving is more dangerous than they previously thought increases the perceived value of \( \mathbb{E}(r) \) and therefore reduces distracted driving. However, providing information beyond a certain threshold is redundant as the the agent either already perceives \( \mathbb{E}(r) \) to be high enough that \( c = 0 \) or values \( c \) so highly that no increases in \( \mathbb{E}(r) \) can further reduce \( c \). Thus, the marginal utility of further information beyond a certain threshold will be 0.

5.3 Unreliable nudge provider model

This function assumes that the agent starts questioning the reliability of the nudge provider if too many nudges are provided. Reasons for this could include paranoia and psychological reactance. The population may lose trust in all government recommendations if some government recommendations are not reliable [?].

Consider the utility function \( u(w, m; i) \) over water \( w \) and all other goods \( m \). \( i \) is the agent’s information about the benefits of drinking water. The optimal level of \( w \) is increasing in \( \mathbb{E}(i) \). Let \( t \) be the number of text message reminders to drink water. As long as the information source is perceived as credible, \( \mathbb{E}(i) \) is
increasing in $t$. Assume that the agent starts to question the reliability of the text message after more than $X$ amount of text messages are received. Let $i_0$ be the agent’s original information about the benefits of drinking water. If the agent does not trust the sender of the messages, $E(i) = i_0$. Assume the agent trusts the sender of the messages unless the sender sends more than $X$ number of text messages a day. If the sender sends more than $X$ messages a day, the sender is perceived as unreliable by the receiver. Thus the benefit of the text messages increases as the number of messages increase until the value $X$ where it drops back to the original values. The value of $X$ may be different for many individuals and the optimal amount of text messages to send would be the summation of the benefit curves of each individual.

6 Criticisms of the theory
One major criticism to the theory that too much nudging can cause a decrease in the desired response derives its basis from the theory of habit formation. Critics could argue that continuous nudging can sustain habit formation which would infer that the effectiveness of the nudge would be directly proportional to the frequency of the repeated nudges. Furthermore, there may also be an argument that suggests that the nudge that causes the habit to be formed may be the most effective and not the first nudge.

A rebuttal to this argument would be that if there is an $n^{th}$ nudge acting as a context cue in the formation of a habit, then any nudge $n^{th} + 1$ would be irrelevant and ineffectiveness in the habit formation. Thus, it may also be possible that such nudges beyond the $n^{th}$ nudge could lead to a negative or opposite effect to what is desired. For example, a smoker who has decided to quit smoking may be nudged to smoke by a non-smoking advertisement that entices the concept of smoking for him again.

Also, while it may be true that a habit is formed only after the $n^{th}$ nudge, it does not mean that the $n^{th}$ was the most important in forming the habit. The first nudge might have had the most influence in causing the habit formation while the $n^{th}$ nudge could have been the one to lead the nudged to automaticity but not been the most influential in the habit formation.

7 Summary of Experimental Data
For this section of the paper, I analysed various research papers and shortlisted relevant ones explaining the significance of their findings in regards to the research question. I found these research papers on Google Scholar using search entries such as "too much nudging", "optimum amount of nudge", "a nudge too much" and similar terms.

In the study Damgaard and Gravert (2018) the researchers studied the effectiveness of 8 reminders in causing their members of the Smithsonian Institution (that funds its activities through charitable donations) to renew their membership in the months just prior to and just after their membership expired. The researchers found the first nudge, that is the first reminder, was highly effective and had a response led to a 23.2% response rate. They also found out that the following reminders were much less effective and had diminishing response rates (see Figure 7).
Gravert and Kurz (2021) also found decreasing effects of repeating a nudge. The researchers initiated a framing nudge that sought to shift restaurant orders from meat dishes towards vegetarian options. The customers in the first week, the results displayed a 21% increase in the probability of selling a vegetarian dish, but the effect diminished to about 6% by Week 3 as seen in Figure 8.

This evidence suggests that nudges become less effective as they are repeated because those exposed to many repetitions of the nudge are those who have intentionally not altered their behavior and therefore are less motivated to do so.

However, both the studies Damgaard and Gravert (2018); Gravert and Kurz (2021) target specific one-time actions such as donating to a charity (Damgaard and Gravert, 2018) and ordering a dish (Gravert and Kurz, 2021). Repeated nudges may be more effective in a sustained long term action such as increased exercise.
Figure 8: Shares of total sales with intervention during weeks 1–3

Di Matola (2020) investigated whether the combination of a green traffic light nudge and a plant-based environment influenced healthy menu options in a restaurant. The green traffic light nudge was a green mark on the menu next to the healthier option (Figure 9) while the plant based restaurant environment had potted plants on the table (Figure 10).

Figure 9: Menus with and without traffic light nudge
The research found that while the two nudges when applied separately led to an increase in the orders of healthy menu options, when working together they negated each other (Figure 11). This supports the initial hypothesis that too much nudging leads to a reduction is desired response. A reason for this could be the excessive nudges caused the nudged to be more conscious of the nudges. Limitations to the experiment were that there were only 82 people as a part of it is rather low for a quantitative study. The menu size was also only 4 options which does not represent actual menu sizes in restaurants. Furthermore, extraneous variables such as the different weather and waiters on different days could have impacted the choice.

Tiefenbeck et al. (2018) used a repeated nudge that provided immediate real-time feedback to those in the shower about their water usage. The study showed a decrease in the water usage post the first nudge but it was relatively stable after. The study did not further investigate the marginal effects of repeating the nudges but the relatively stable and even slightly increasing graph (Figure 12) suggest that the repeated nudges may eventually have a
decreasing effect as well. Another point to note in this study is that the participants were self-selecting and hence may be more environmentally concerned than a simple random sample.

![Figure 12: The Impact of Real-Time Feedback on Energy (and Water) Consumption](image)

Vitek and Syed Shah (2019) studied the effectiveness of nudges as a tool to prevent phishing. The research found out that using a white-list approach that reduces the number of nudges over time was the most effective as it counter-acted desensitization the best. A white-list approach to data validation in this context is when there is a list of administrator approved email addresses that don’t receive nudges indicating the possibility of phishing. Once the email address is approved subsequent emails from the same address no longer warrant a nudge to warn for phishing. Thus, the nudges decrease over time. The conclusion of a white-list approach to be the most effective is in line with the hypothesis that too much nudging can cause a decrease in the desired response. In this case, the nudged are desensitized to the nudges over time. As the reduction in the frequency of the nudges over time matches the desensitization, the importance of the nudge to the nudged and thus its impact is relatively constant.

8 Conclusion

The initial research question was "can too much nudging cause a decrease in the desired response?" and according to my research there is sufficient evidence to assume so. All 5 of the experiments I analysed had data that supported the hypothesis. However, each of the experiments did not directly answer the research question and had limitations. Furthermore, through the course of the research I found that nudges have drastically different results depending on the activity they influence, the type of nudge, the target audience to be nudged and various external factors as well. If the activity was a long-term continuous activity or a one time action; if the target audience was conscious of the nudge or not; if the nudge changed in frequency, intensity or duration are all relevant to the effect of the nudge. There are many obvious as well as subtle factors to be considered when dealing with a psychological concept such as a nudge. However, we could assume that if all other things remained constant, there would be a point where a nudge would be too much.
9 Further Research

The best suggestion for further research would be to conduct an long-term and large-scale experiment that exactly pinpoints a Laffer Curve type of curve for the relationship between repeated nudge frequency and nudge effectiveness.

An example of a model experiment would have a certain number of participants sign up for text messages that reminded them to drink water. Stratified samples would be used to divide participants in various groups for different frequencies of text messaging as it would account for variances in water intake, phone usage, attention span and external factor variance amongst people of different ages, gender and other differentiators. 10 or more groups would be ideal. The variance in the frequency of messages could be from one message weekly to 20 in a day. The experiment would first last for 2 weeks without any text message reminders to measure their baseline water intake amount of each participant. Then the experiment would last for 4 more weeks to measure the change in water intake after the repeated nudges. Through this data collected a more accurate curve could be modeled to illustrate the effects of nudging too much.

The experiment could also repeated with different variances in reminder frequencies, different duration of the text message section of the experiment and changes to the action influenced. For example, weight loss or amount of physical exercise could be measured. However, some of the limitations to the experiment would include the influence of framing of the reminder itself and its impact on people. Furthermore, the experiment cannot be generalised to a "nudge" as an overarching statement because of the specificity. Thus, only with a broad range of experiments looking into this niche of nudging can we truly determine the ill effects of nudging too much.

Acknowledgments

I would like to thank my advisors Rohen Shah and Philip Liang for their guidance in this project.

References


