

# Nudging Towards Inclusivity: The Ethicality of Policies for the Promotion of Inclusion

*By Manuj Kant, Las Positas College  
Mentor: Jeremiah Bodnar*

## Abstract

There can sometimes be weariness surrounding coordinated efforts to induce inclusion. Detractors of certain such efforts may cite coercion or a loss of autonomy as a concern. In their book *Nudge*, Thaler and Sunstein presented the idea of the “nudge.” Nudges change the context in which a decision is made to increase the likelihood that a certain choice is made by the decision maker. However, nudges are designed to be easy to avoid. Thaler and Sunstein laid out criteria such that nudges can remain ethical. In this paper, I show that certain nudges that greatly increased inclusion in their respective communities do follow these criteria. In fact, subjected even to the relatively tighter constraints imposed by a prominent critic of Thaler and Sunstein’s philosophy, Wilkinson, these nudges remain unscathed. Thus, inclusion-promoting nudges are not necessarily unethical from the standpoint of manipulation or autonomy.

## Introduction

In their book *Nudge: Improving Decisions About Health, Wealth, and Happiness*, Thaler and Sunstein presented the idea of “nudges.” They defined nudges 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” (Thaler and Sunstein, 2021, p. 8). The term choice architecture refers to the different ways a choice can be presented. Additionally, Thaler and Sunstein required that nudges be “easy and cheap to avoid” (Thaler and Sunstein, 2021, p. 8). Thaler and Sunstein claimed nudges are a form of “libertarian paternalism.” They are libertarian because nudges preserve people’s freedom of choice and paternalistic because they steer people to make choices that are good for them.

There have been numerous instances in which nudges effectively altered human behavior. In one well-documented case, the choice architecture of a company’s retirement plan was changed from an opt-in to an opt-out design. This means that the default option was changed such that employees were automatically opted into the plan, but could choose to opt-out if they wanted to. This nudge caused employee participation in the retirement plans to increase significantly (Thaler and Sunstein, 2021, pp. 179-197). Another classic example of a nudge decreased “spillage” at Amsterdam’s Schipol Airport’s men’s toilets by 80 percent. By drawing flies on the insides of the toilets, men “spilled” less as they were nudged to aim at the fly. (Thaler and Sunstein, 2021, pp. 4-5).

A fertile ground for the application of a libertarian paternalistic methodology is in the design of policies that seek to increase inclusion in a society. While the inclusion of people of diverse backgrounds is widely regarded as a societal good, policies that attempt to achieve this can sometimes be perceived as being coercive. In their study centered around effective diversity statements for medical centers, *Promises and Pitfalls of Diversity Statements: Proceed with Caution*, Carnes et al. acknowledged this unfortunate reality, stating “statements that emphasize autonomy avoid the backlash that can occur when organizational members feel coerced into adopting pro-diversity actions” (Carnes et al., 2019). Nudges designed to promote inclusivity, that are not manipulative, and thus, do not infringe upon the autonomy of the targets, could be a suitable alternative to coercive measures to improve diversity.

While Thaler and Sunstein believed that nudges could have positive effects, they were also aware that some nudges may be manipulative or used for malevolent purposes. They proposed two criteria—publicity and transparency—to prevent these so-called “evil nudges.” The publicity principle requires that nudges only be used when a choice architect is willing to publicly defend their use (Thaler and Sunstein, 2008, p. 244). Given that the criterion is met, it is less likely for the nudge to be harmful due to the fact that it meets a societal consensus. The principle of transparency requires that a nudge not be “invisible and thus impossible to monitor” (Thaler and Sunstein, 2008, p. 246). Thus, a given nudge should be relatively open in the sense that those being nudged should be cognizant of the fact that they are being influenced in a certain way. Thaler and Sunstein give the example of subliminal messaging as a nudge that is not transparent, and thus evil. They argued that nudges following these two criteria can be considered ethical. They claimed that since some form of choice architecture must exist in general, nudging is inevitable. Thus, there is no reason for choice architects to not apply nudges that will benefit people. (Thaler and Sunstein, 2008, p. 237).

## Wilkinson’s Criticisms

In his well-cited article *Nudging and Manipulation*, T. M. Wilkinson argued that Thaler and Sunstein’s justification of nudges due to their inevitability is not strong. Wilkinson pointed out that choice architecture can be created randomly rather than with the intention of steering people to pick a certain choice. This would ensure that people are not being controlled by the choice architect. (Wilkinson, 2013, p. 343). Thus, it is not inevitable that people will be nudged by the architect.

Additionally, Wilkinson critiqued Thaler and Sunstein’s criteria for a nudge to not be manipulative. However, these criteria can be refined to address Wilkinson’s concerns. Regarding the publicity criterion, according to Wilkinson, Thaler and Sunstein failed to explain what it means for a choice architect to be “able” to defend a nudge. Wilkinson brought up the issue that a choice architect may be willing to publicly defend a nudge that is harmful (Wilkinson, 2013, p. 344). To account for this criticism, I claim that this criterion can reasonably be interpreted as meaning that the architect should be able to win a public vote on the net good of their implementation of the nudge. This definition clearly goes well beyond the architect simply being willing to defend their nudge and forces the nudge to be acceptable to society. In addition to this criticism, regarding the transparency criterion, Wilkinson questioned who Thaler and Sunstein were suggesting should monitor the nudge. For the purpose of this analysis, I interpret Thaler and Sunstein’s transparency requirement to mean that it should be practically possible for the *target* of the nudge to notice they are being nudged. This is the tightest possible requirement since the purpose of transparency is for those affected by an action to understand that it is being taken.

Furthermore, Wilkinson introduced a couple of additional criteria for determining the ethicality of a nudge. The first of these has to do with the nudge having a “genuine escape clause” (Wilkinson, 2013, p. 354). The root of the necessity for this requirement is Wilkinson’s definition of manipulation as “intentionally and successfully influencing someone using methods that pervert choice” (Wilkinson, 2013, p. 347). Manipulation is wrong because it violates the autonomy of people by “subverting and insulting their decision-making powers” (Wilkinson, 2013, p. 351). By Wilkinson’s definition, manipulation requires a manipulator to intentionally make someone do something. Nudges are intentionally designed to steer people toward certain choices. However, they do not *prevent* people from choosing options they are not being nudged towards. Wilkinson acknowledged that, according to Thaler and Sunstein, when influenced by an ethical nudge, people should be able to pick other options if they are being nudged to pick an option they do not like (Wilkinson, 2013, p. 351). Wilkinson argued that nudges are only manipulative when the nudger has manipulative intentions, and a nudger does not have manipulative intentions if a nudge has a genuine escape clause. He further clarified that a nudge has a genuine escape clause when the “nudger sincerely wants the targets not to act in the nudged way if the

nudging is unsuitable for them” (Wilkinson, 2013, p. 354). Wilkinson’s second additional criterion was that nudges should not pervert choice. He used a piecemeal approach to determine which nudges pervert choice as there is not a good general account. According to Wilkinson, nudges that involve lying or subliminal messaging likely pervert choice. However, nudges that simply present information using knowledge from social psychology and behavioral economics to nudge people toward a certain choice may not be manipulative. These nudges provide decision-makers with a better understanding of the facts to make their decision.

In what follows, three inclusion-promoting policies are subjected to a test of ethicality under Thaler and Sunstein’s criteria for ethical nudging, along with the refinements and additions of Wilkinson.

## **Nudging to Improve the Diversity of Self-Assembled Teams**

Crowdsourcing is a resource companies have recently begun employing to outsource innovation.

Crowdsourcing teams are self-assembling teams of people who collaborate to create products and innovative solutions. However, due to the self-assembling nature of these teams, they can lack diversity due to the homophilic biases of team members (Vinella et al., 2022, p. 265). This is harmful because diversity has been shown to improve the quality of crowdsourced teams. Studies have shown that Wikipedia teams composed of people with opposing political views perform better than groups with similar views (Vinella et al., 2022, p. 266).

In their study, *The Impact of Digital Nudging Techniques on the Formation of Self-Assembled Crowd Project Teams*, Vanilla et al. determine whether nudging can improve the diversity of crowdsourcing teams. They tried two nudging techniques on participants that were designed to improve the diversity of teams. The first nudge displayed a personalized diversity score and diversity recommendations based on the team a participant had picked. The second used diversity priming through “exposure to counter-stereotypes and the all-inclusive multiculturalism approach” (Vinella et al., 2022, p. 267). Exposure to counter-stereotypes was done by displaying positive images of people from minority groups. The all-inclusive multiculturalism approach attempts to promote inclusivity by “explicitly mentioning both majority and minority groups” (Vinella et al., 2022, p. 267). This approach ensures that the importance of minorities is considered while people from majority groups do not feel excluded.

The study found that the first approach, in which nudges that displayed personalized diversity information about the chosen group, increased the diversity of groups for crowdsourcing projects. However, nudges that used diversity priming, the second approach, had the opposite effect and strengthened homophilic biases. Generally, there was a tendency for groups to pick people that were the same gender as them (Vinella et al., 2022, 274). This conclusion shows that while some nudges can be effective in increasing diversity, it is important to be careful in considering which nudges to use, as some nudges may have an adverse effect on diversity.

### ***Analysis of Nudge***

Since diversity priming was unsuccessful at increasing inclusion, for the purpose of this analysis, consider only the first approach from the above example. This nudge meets the publicity and transparency criteria Thaler and Sunstein use to differentiate good nudges from “evil nudges.” Consider first the publicity criteria. Is the provision of information about the diversity of a group a defensible societal position? While specific policies about the degree to which considerations of diversity should influence team-building may be contentious, the simple pointing out of the level of diversity in a team is relatively socially acceptable. A survey by Pew Research revealed that 75% of Americans believe “it is very/somewhat important for companies and organizations to promote racial and ethnic diversity in their workplace” (Horowitz, 2019). At the very least, it is not a measure that the vast majority of society would strongly oppose. Then, consider the transparency

requirement. On this criterion, the given nudge passes with flying colors. There is nothing subliminal or sneaky about giving a person metrics regarding the diversity of their team. It must be apparent to the target of the nudge that they are being influenced not only to consider the level of diversity of their team but, in particular, to consider increasing that level, for there is no other reasonable motive behind this nudge.

Then, consider Wilkinson's additional criteria for the ethicality of nudges. The first criterion requires the nudge to have a genuine escape clause. Here, if the target does not want to change their team on the basis of the presented information, they can simply ignore the information. The team selection software was not designed to have obstacles that prevented participants from creating homogenous teams. Thus, it can be reasonably concluded that the choice architect wanted those who found the nudge to be unsuitable to act against the direction in which they were nudged, so there is a genuine escape clause. Wilkinson's second criterion asserts that the target's decision-making should not be perverted by the nudge. While adherence to this criterion is difficult to judge, Wilkinson indicated that the presentation of information based on social psychology can be accepted as satisfying it. Since this nudge purely involves making such information available to the target, it is reasonable to conclude that it does not pervert the decision-making process of the target. Thus, by all of the aforementioned criteria, the presentation of a diversity score and other diversity-related information in crowdsourced team-building can be judged to be ethical.

## **Opt-out Choice Framing and Female Participation in Competitions**

Women tend to occupy a lower proportion of high-level positions compared to men (He et al., 2021, p. 1). Some methods, such as unconscious bias training, have been employed to attempt to close this gender gap. However, these methods have not resulted in a significant improvement. In their article, Opt-out choice framing attenuates gender differences in the decision to compete in the laboratory and in the field, He et al. tested whether a nudge that modifies the choice architecture of the decision to participate in a competition (or not) from opt-in to opt-out could encourage women to compete against their peers. In the opt-out choice architecture, applicants participated in the competition by default. They had to actively choose to opt-out if they so desired. By contrast, in the opt-in choice architecture, by default, applicants did not participate, and had to actively choose to participate if they were interested in doing so. They pointed out that the results from this experiment could have implications for reducing the gender gap in high-level positions, as the decision to participate in a competition is analogous to the decision to apply for such a position.

In the study, participants were given the task of adding five two-digit numbers together as many times as possible in five minutes (He et al., 2021, pp. 1 - 2). In stage one of the study, participants performed the task, receiving 50 cents for each correct answer they gave. In stage two, participants performed the task again. However, they received a competition-based reward this time. This reward was determined by comparing their score to three other randomly selected participants' scores. In this group of four participants, the participant with the highest score would receive two dollars per correct answer. The other three participants would receive nothing. In stage three of the study, participants were asked to repeat the task, but were given a choice between the competitive compensation scheme from stage two or the standard scheme from stage one. Participants would make their selection by checking either one of the boxes on a software developed by the researchers. Participants were randomly assigned either an opt-in choice architecture or an opt-out choice architecture for this decision. If a participant was assigned the opt-in architecture, the box for the uncompetitive compensation scheme was automatically ticked. Likewise, if a participant was assigned the opt-out architecture, the box for the competitive compensation scheme was automatically ticked. In both cases, participants could manually choose to tick the box for the other compensation scheme if they wished to.

The opt-out choice architecture increased the number of women who participated in the competitive compensation scheme. Of the participants who were given the opt-in condition, 72.48% of men and 46.72% of

women chose the competitive compensation scheme (He et al., 2021, p. 2). However, of the participants who were given the opt-out condition, 76.42% of men and 75.38% of women chose the competitive compensation scheme. Thus, the nudge successfully promoted women to compete, and a similar nudge could be employed to improve the number of women who apply for high-level positions.

### ***Analysis of Nudge***

This nudge meets Thaler and Sunstein's publicity and transparency principles. At large, people support promoting women to apply for high-level positions. A survey by Pew Research revealed that 59% of adults believe there are too few women in top executive business positions (Horowitz et al., 2018). The nudge also meets Thaler and Sunstein's transparency requirement. Consider the targets of this nudge--the people (in particular, the women) who might otherwise (under a different choice architecture) have chosen to not compete. This means that the targets must have had some inclination to abstain from competition. Given this, they must have found it somewhat jarring that, under the opt-out architecture, the competitive option was the one selected by default. Thus, the targets would have noticed that they were being nudged, indicating the transparency of the nudge.

This nudge also meets Wilkinson's criteria for a nudge to be ethical. Wilkinson's first criterion requires the nudge to have a genuine escape clause. As stated previously, participants were not forced to choose the competitive or uncompetitive option. The study was designed such that participants had the ability to select the other option if the automatically selected option was not the one they preferred. Participants could change their choice from the default option without facing any obstacles. In order to do this, they simply had to click on the box for the other option (He et al., 2021, p. 2). Thus, it can be reasonably concluded that the nudgers wanted participants who genuinely did not want to compete to select the uncompetitive option. Wilkinson's second criterion requires that nudges do not pervert the target's decision-making. While this criterion is difficult to judge, changing to an opt-out choice architecture does not seem to fit with the deceitful and tempting nudges that Wilkinson claims pervert the decision-making of the target (Wilkinson, 2013, p. 348). Since it adheres to the aforementioned criteria, changing the choice architecture of the decision to pick between a competitive versus an uncompetitive option from opt-in to opt-out can be judged to be ethical.

### **Nudging to Increase Socioeconomic Diversity in Undergraduate Economics**

There is a lack of diversity in economics as minority groups such as Black, Hispanic, and Native American people are underrepresented (Pugatch and Schroeder, 2021, p. 2). Without the valuable perspectives of people from these groups, important economic questions could go unexplored. In Pugatch and Schroeder's study, *A Simple Nudge Increases Socioeconomic Diversity in Undergraduate Economics*, the authors attempted to determine whether sending emails containing information about economics majors promoted diversity among students pursuing an economics major at Oregon State University.

In their study, emails were sent to 2,200 students taking an economics course at the university (Pugatch and Schroeder, 2021, p. 2). Some randomly selected students were sent a video that showed them rewarding careers in the field of economics. Another group of randomly selected students was sent a "financial returns" message that compared the financial opportunities of an Economics major to those of students in other majors. These two groups of students were compared to students who did not receive an email or received an email containing only basic information about the Economics major.

The email with basic information had the largest positive impact on diversity in the major (Pugatch and Schroeder, 2021, p. 2). It increased the probability that a student from an underrepresented minority group

major in economics by five percent. This was enough to “more than reverse the gap in majoring in Economics for students of this demographic group” (Pugatch and Schroeder, 2021, p. 2).

### ***Analysis of Nudge***

In this analysis, I will focus on the email containing basic information as it was the nudge that had the most significant positive impact. This nudge meets Thaler and Sunstein’s publicity and transparency criteria. One can reasonably assume that a vast proportion of the population has no issue with sending students basic information about majoring in economics. Thus, the creators of this nudge would be able to publicly defend it. The nudge also meets Thaler and Sunstein’s transparency criterion. The nudge is extremely visible as targets of the nudge intentionally click on and read the message. Additionally, targets of this nudge can be expected to understand that they were sent the message to increase their interest in pursuing the economics major since emails of this type (across many departments at many colleges) are relatively common and typically sent for this purpose.

The nudge also meets Wilkinson’s criteria for a nudge to be ethical. Wilkinson’s first criterion requires the nudge to have a genuine escape clause. The targets of this nudge could easily choose to major in any field after reading the message. Additionally, it is doubtful that the implementers of such a nudge (likely the economics department at a particular college) would want students who were completely uninterested in economics to major in the field simply because of an informational email.

Thus, the nudge has a genuine escape clause. Wilkinson’s second criterion requires that nudges not pervert the target’s decision-making. However, he asserts that informational nudges do not do this. Thus, based on the aforementioned criteria, we can conclude that emailing students basic information about economics is ethical.

### **Conclusion**

In this article, I used Thaler and Sunstein’s and Wilkinson’s criteria for ethical nudges to analyze three nudges that could greatly increase inclusion in crowdsourced teams, high-level positions, and economics departments. Thaler and Sunstein required that nudges be publicly defensible and transparent, and Wilkinson required that nudges have a genuine escape clause and not pervert the choice of the target. When analyzed through the lens of these criteria, the considered nudges are revealed to be non-manipulative and ethical. This shows that while some inclusivity-promoting policies may receive backlash due to their perceived coerciveness, well-architected nudges for the promotion of inclusion do not necessarily infringe upon autonomy.

## **References**

- Carnes, M., Fine, E., & Sheridan, J. (2019). Promises and pitfalls of diversity statements. *Academic Medicine*, 94(1), 20–24. <https://doi.org/10.1097/acm.0000000000002388>
- He, J. C., Kang, S. K., & Lacetera, N. (2021). Opt-out choice framing attenuates gender differences in the decision to compete in the laboratory and in the field. *Proceedings of the National Academy of Sciences*, 118(42). <https://doi.org/10.1073/pnas.2108337118>
- Horowitz, J. M. (2019, May 8). *Americans see advantages and challenges in country's growing racial and ethnic diversity*. Pew Research Center's Social & Demographic Trends Project. Retrieved February 14, 2023, from <https://www.pewresearch.org/social-trends/2019/05/08/americans-see-advantages-and-challenges-in-countrys-growing-racial-and-ethnic-diversity/>
- Horowitz, J. M., Igielnik, R., & Parker, K. (2018, September 20). *Women and leadership 2018*. Pew Research Center's Social & Demographic Trends Project. Retrieved February 14, 2023, from <https://www.pewresearch.org/social-trends/2018/09/20/women-and-leadership-2018/>
- Pugatch, T., & Schroeder, E. (2021). A simple nudge increases socioeconomic diversity in undergraduate economics. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3860599>
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin Books.
- Thaler, R. H., & Sunstein, C. R. (2021). *Nudge: The final edition*. Penguin Books, an imprint of Penguin Random House LLC.
- Vinella, F., Mosch, R., Lykourantzou, I., & Masthoff, J. (2022). The impact of digital nudging techniques on the formation of self-assembled crowd project teams. *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. <https://doi.org/10.1145/3503252.3531298>
- Wilkinson, T. M. (2013). Nudging and manipulation. *Political Studies*, 61(2), 341–355. <https://doi.org/10.1111/j.1467-9248.2012.00974.x>