

Do Not Commit Evil Even if It is Small: Psychological Mechanisms of the Slippery Slope Effect and Prevention Strategies

Authors: Ou Honglei, Sun Binghai, Zhang Wenhai, Weijian Li, Li Weijian

Date: 2019-12-25T00:00:00+00:00

Abstract

The slippery slope effect refers to the tendency for an individual's level of immorality to gradually intensify over time, meaning that immoral behavior exhibits repetitiveness and progressive escalation. It is widely present both at the individual level and in interpersonal interaction processes, affecting individuals' normal social interactions and even hindering the harmonious and healthy development of society. Research approaches to the slippery slope effect include experimenter manipulation methods and spontaneous deception methods; psychological mechanisms encompass moral disengagement, implicit bias, self-dehumanization, among others; corresponding intervention measures involve enhancing prevention-focused motivation and shaping positive descriptive social norms. Future research could further investigate the applicable scope of the slippery slope effect, neural intervention approaches, and other aspects from the perspectives of psychological processes and physiological mechanisms.

Full Text

The Slippery Slope Effect: Psychological Mechanisms and Prevention Strategies

OU Honglei¹, SUN Binghai^{1,2}, ZHANG Wenhai³, LI Weijian¹

¹College of Teacher Education, Zhejiang Normal University;

²Tin Ka Ping Moral Education Research Center, Zhejiang Normal University, Jinhua 321004, China

³Mental Health Education Center, Yancheng Institute of Technology, Yancheng 224003, China

Abstract

The slippery slope effect refers to the tendency for individuals' degree of im-

morality to intensify gradually over time, meaning that unethical behavior exhibits both repetitive and progressive characteristics. This phenomenon exists widely at both the individual level and in interpersonal interactions, affecting normal social communication and even hindering the harmonious and healthy development of society. Research methods for investigating the slippery slope effect include experimenter manipulation and spontaneous deception, while its psychological mechanisms encompass moral disengagement, implicit biases, and self-dehumanization. Corresponding interventions involve enhancing prevention-focused motivation and shaping positive descriptive social norms. Future research should further explore the applicable scope of the slippery slope effect and its neural intervention methods from the perspectives of psychological processes and physiological mechanisms.

Keywords: slippery slope; moral disengagement; implicit biases; self-dehumanization; prevention focus; descriptive social norms

During the Three Kingdoms period, Liu Bei admonished his descendants in his will: “Do not commit evil deeds, however small,” urging his son Liu Shan to cultivate virtue and avoid minor transgressions, for accumulated small evils “could bring disorder to the state.” In the Northern and Southern Dynasties, the *Book of the Later Han* proposed the concept of “preventing gradual deterioration,” arguing that erroneous thoughts or behaviors should be stopped promptly when they first emerge to prevent their further development. Since ancient times, such ideas of nipping problems in the bud have been passed down, and the notion of preventing the escalation of unethical behavior has long been deeply rooted in people’s minds. Western researchers have termed this “tendency for unethical behavior to repeat and intensify” as the **slippery slope effect**, whereby after an individual commits an unethical act, their moral level plummets like a snowball rolling down a slope, becoming difficult to restrain.

As early as 1975, the slippery slope effect was manifested in Milgram’s well-known obedience experiment: under punitive instructions, participants gradually administered increasingly stronger electric shocks to innocent individuals, eventually reaching lethal levels (Milgram, 1975). At that time, research focused solely on authority obedience, and no one recognized the implicit slippery slope effect. Later, researchers argued that one key feature of Milgram’s experimental situation that led to high obedience rates was the incremental nature of shock intensity (Burger, 2009). Had the shock instructions suddenly jumped from a weak voltage to a lethal level, perhaps more participants would have refused to carry out the harmful behavior.

In recent years, the slippery slope effect has entered the stage of empirical research. Researchers both domestically and internationally have conducted in-depth studies, continuously making new progress in exploring its psychological and physiological mechanisms as well as preventive measures (陈银飞, 2011; Garrett, Lazzaro, Ariely, & Sharot, 2016; Tang et al., 2019; Welsh, Ordonez, Snyder,

& Christian, 2015; Yin & Weber, 2019). This paper attempts to provide a systematic and comprehensive synthesis of the slippery slope effect, advocating for a shift in moral research from single time points to a dynamic developmental perspective. We first outline the slippery slope effect, then summarize its research methods, causes, and intervention strategies, and finally propose directions for future research, hoping to provide reference for related studies and applications.

1. Concept and Research Overview of the Slippery Slope Effect

The slippery slope effect refers to the tendency for individuals' unethical behavior to be repetitive, with the degree of immorality gradually intensifying over time. Similar to the process of small evils accumulating into major ones, it emphasizes the repetitive and progressive nature of unethical behavior (Baron, Zhao & Miao, 2015; Lafollette, 2005; Welsh et al., 2015).

Early discussions of the slippery slope effect were primarily theoretical reviews based on life experience, social issues, and previous viewpoints. Chinese adages such as "A small hole not patched will become a big hole that causes suffering" and "Prevent gradual deterioration" remind people to be vigilant of the first signs of wrongdoing. Foreign scholar Schauer concluded from numerous real legal cases that a seemingly innocuous act that initially appears not to violate moral standards may subsequently lead to a chain of increasingly harmful similar events (Schauer, 1985). Lafollette argued that if a less unethical act X occurs, after a series of minute changes, a highly unethical act Y is likely to occur (Lafollette, 2005). These perspectives all suggest that the slippery slope effect is like boiling a frog in warm water—people fail to notice the slow changes in their own unethical behavior, easily leading to catastrophe.

Recent research has introduced the slippery slope effect into the empirical domain. In domestic research, Zheng Qinqin (2009) used event history analysis with longitudinal data from 2000-2007 to examine information disclosure violations of listed companies, finding a high correlation between companies' previous unethical behavior and the incidence of subsequent unethical behavior. Chen Yinfei and Mao Ning (2011) employed experimental methods to study the temporal changes in ethical judgment biases of corporate decision-makers, showing that unethical behavior gradually worsens rather than suddenly intensifying. Since then, domestic research in this area has not made much progress. Meanwhile, foreign studies indicate that individuals are more likely to engage in unethical behavior when rewards increase gradually rather than suddenly (Welsh et al., 2015), and to accept others' unethical behavior (Gino & Bazerman, 2009). The nature of unethical behavior becomes more serious over time (Engelmann & Fehr, 2016). Recently, domestic scholars have begun to follow up, using near-infrared spectroscopy and functional magnetic resonance imaging to further explore the neural mechanisms of the slippery slope effect (Tang et al., 2019; Yin et al., 2019).

In summary, previous research shows that the slippery slope effect exists at both the individual level and the interpersonal interaction level. At the individual level, the moral slide occurs within the individual without affecting others. For example, Welsh et al. demonstrated the existence of the individual-level slippery slope effect, showing that participants' deception amounts gradually increased over time (Welsh et al., 2015). At the interpersonal interaction level, the slippery slope effect easily emerges in interactions between people and between groups (Ashforth & Anand, 2003; Courtois & Gendron, 2017). For instance, Gino et al.'s research revealed the interpersonal-level slippery slope effect: when two participants were collaborators and one engaged in increasingly serious unethical behavior, the other gradually became involved as well (Gino et al., 2009). Volokh proposed the intergroup-level slippery slope effect (Volokh, 2003), arguing that it easily occurs between groups. Group polarization phenomena, where certain deviant viewpoints or behaviors of some group members may eventually lead to atrocities at the group level, exemplify this (Moscovici & Zavalloni, 1969).

When discussing the slippery slope effect, we must guard against the “slippery slope fallacy.” This fallacy refers to the unreasonable use of a chain of causal relationships, transforming all “possibilities” into “necessities” to draw conclusions that some wish to prove. The logic of the slippery slope fallacy is: If A occurs, then B will certainly occur, leading to C...and eventually causing serious consequence Z. Therefore, A must absolutely not be allowed. This “if such-and-such event occurs, then some bad thing will follow” argument is overly extreme and one-sided. In understanding the slippery slope effect, we must avoid sweeping generalizations, recognizing that it represents only a tendency, not a necessity.

In conclusion, unethical behavior has a tendency to change gradually over time, a process that pushes people's unethical actions toward increasingly corrupt states—this is called the slippery slope effect. Based on these characteristics, the following section introduces research methods for studying this phenomenon.

2. Research Methods for the Slippery Slope Effect

Based on its repetitive and progressive nature, research methods for the slippery slope effect have unique characteristics: with slight modifications to classic moral paradigms, the slippery slope effect in single-person or interpersonal interaction modes can be effectively simulated. Currently, two major categories of adapted methods have been used: first, the **experimenter manipulation method**, which requires researchers to create both slippery slope and abrupt change conditions; and second, the **spontaneous deception method**, which requires no additional manipulation, only extending the number of rounds to observe spontaneous changes in participants' behavior.

2.1 Experimenter Manipulation Method

This method emerged in early research on the slippery slope effect (陈银飞, 2011). It typically manipulates task rewards or difficulty by creating “gradual change”

and “abrupt change” conditions. The former represents the slippery slope condition, where experimental rewards or difficulty stimuli change slowly and minimally; the latter represents the abrupt change condition, where rewards or difficulty stimuli change suddenly and substantially. Both conditions have the same number of experimental rounds. By comparing differences in unethical behavior in the final round between the two conditions, researchers can demonstrate the existence of the slippery slope effect. Below are two examples of this method.

Task Reward Manipulation. The paper-and-pencil matrix task, developed by Mazar, includes a test sheet with multiple number matrices, an answer sheet, and an envelope containing participant payment. Participants are told they will receive a certain amount of money for each matrix solved. They must circle two numbers in each matrix that add up to 10, then write down the number of matrices they solved correctly on the answer sheet. At the end of the experiment, all participants are asked to throw their test sheets into a box, submit only their answer sheets, and take the appropriate payment from the envelope, returning the remainder (Mazar, Amir, & Ariely, 2008). This procedure makes participants believe their true performance will not be known, though researchers can determine whether they lied by matching answer sheets with test sheets. In slippery slope research, researchers extend the task to three rounds, with reward amounts set to increase gradually or suddenly. In the slippery slope condition, rewards per matrix are \$0.25, \$1.00, and \$2.50 across the three rounds. In the abrupt reward condition, there is no reward in the first two rounds, but participants receive \$2.50 per matrix in the third round. If cheating rates in the third round are significantly higher in the slippery slope condition than in the abrupt condition, the slippery slope effect is confirmed (Welsh et al., 2015). Similar adaptations have been made to perceptual tasks and the corruption game (Kobis, van Prooijen, Righetti, & Van Lange, 2017).

Task Difficulty Manipulation. The mental math task requires participants to solve ten addition and subtraction problems. Researchers inform participants that they must press the space bar when the task begins; otherwise, the answer will appear on the screen after a few seconds. The computer automatically checks whether participants cheat by taking shortcuts (Hippel, Lakin, & Shakarchi, 2005). In slippery slope research, problem difficulty continuously increases in the slippery slope condition, while remaining low until suddenly increasing in the final round in the abrupt condition. If cheating in the final round is significantly higher in the slippery slope condition, the effect is confirmed.

2.2 Spontaneous Deception Method

In this type of slippery slope research, researchers do not manipulate additional variables but simply extend the number of rounds in classic paradigms to observe participants’ spontaneous deception. Existing adaptations are based on classic moral paradigms, including single-player coin-tossing and dice-rolling tasks without interaction partners, and two-party tasks with interaction part-

ners (Fischbacher & Follmi-Heusi, 2013; Weisel & Shalvi, 2015). Single-player spontaneous deception tasks are relatively easy to understand and manipulate—for example, repeatedly tossing coins or rolling dice oneself, with different outcomes corresponding to different monetary rewards, to examine participants’ lying trends (Henderson, Thoelen, Nadler, Barraza, & Nave, 2018). Two-party spontaneous deception research methods are more complex but similarly involve adapting classic paradigms to multiple rounds. Below is a frequently cited example of the two-party task.

The two-party task involves two roles: an advisor and an estimator, who must cooperate to solve a task estimating the number of coins in a jar. The advisor sees a clear picture of the jar with a range of coin numbers, while the estimator sees a blurry picture without a range. During the formal experiment, the advisor must provide monetary advice to the estimator based on available information; the estimator can accept or reject the advice and submit a final estimate. In reality, the advisor is the real participant while the estimator is a confederate. Participants are randomly assigned to one of three conditions for unethical behavior: “self-benefiting-other-benefiting,” “self-benefiting-other-harming,” or “self-harming-other-benefiting.” In the “self-benefiting-other-benefiting” condition, the more the participant overestimates, the more both parties earn; in the “self-benefiting-other-harming” condition, the more the participant overestimates, the more they earn and the less the confederate earns; the “self-harming-other-benefiting” condition is the opposite. Additionally, all participants have an extra earning opportunity: after the experiment, they are asked to make an accurate estimate of the coin amount in a jar shown earlier in the task and receive rewards based on accuracy (Cain, Loewenstein, & Moore, 2005), allowing researchers to determine whether participants lied during the experiment. The research focuses primarily on the “self-benefiting-other-harming” condition, finding that in this “harm your partner for more rewards” situation, participants’ lying levels show an upward trend as game rounds increase.

In fact, most slippery slope research based on classic moral paradigms can approach the phenomenon from these two angles. For example, researchers could also set up slippery slope and cliff conditions in the two-party task, where the confederate advisor provides advice that deviates from the actual amount either gradually or suddenly, and the real participant estimator chooses whether to accept or reject the advice, examining participants’ acceptance of gradually overestimated advice values across different patterns (Gino et al., 2009).

3. Causes of the Slippery Slope Effect

Over the past decades, scholars have continuously explored the internal causes of the slippery slope effect, making considerable progress. Summarizing previous research, the psychological causes can be divided into two perspectives: the bounded ethicality perspective and the adaptation perspective, each with different emphases. The bounded ethicality perspective frames the slippery slope effect as a loss of reason, viewing it as a result of diminished rationality; the

adaptation perspective leans toward emotion, considering it a result of sensory and emotional adaptation. Below we elaborate on several psychological mechanisms of the slippery slope effect.

3.1 Rationality: The Bounded Ethicality Perspective

Simon (1956) proposed the concept of bounded rationality, suggesting that human rationality lies between complete rationality and irrationality. When facing moral conflicts, individuals' rationality is similarly constrained and becomes depleted—this is the **bounded ethicality** theory (Banaji, Bazerman, & Dolly, 2003; Rees, Tenbrunsel, & Bazerman, 2019). From this perspective, unethical behavior is irrational, intuitive, and automatic, leading actors to be in an unconscious and irrational state when committing unethical acts. This state is the core factor causing the slippery slope effect, as individuals do not perceive the moral nature of their behavior during the process. According to domestic and foreign scholars, causes of the slippery slope effect from this perspective include moral disengagement, ego depletion, and implicit biases.

Moral Disengagement. Society demands that people follow moral values, so individuals must constantly regulate the conflict between interests and their own moral standards. They maintain a positive moral self-concept while allowing themselves to commit minor moral transgressions (Moore, Detert, Trevino, Baker, & Mayer, 2012). In other words, individuals tend to rationalize minor moral errors, gaining benefits without being forced to negatively update their moral self-concept (Gino et al., 2009; Mazar et al., 2008; Shalvi, Handgraaf, & De Dreu, 2011). Over time, more and more moral transgressions are incorporated into the moral self-concept (Tenbrunsel & Messick, 2004). Eventually, corruption becomes the norm (Ashforth et al., 2003).

This internal psychological process is called **moral disengagement**, through which people develop motivated forgetting of their own unethical behavior (Shu, Gino, & Bazerman, 2011). Previous research has shown that moral disengagement can explain people's tendency to engage in various unethical behaviors, such as aggression (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996), deception (Detert, Trevino, & Sweitzer, 2008), embezzlement (Moore, 2008), and lying (Moore et al., 2012). This evidence indicates that as moral disengagement increases, the range of actions individuals allow themselves to take expands, making them particularly prone to “sliding further down the unethical slope” (Tenbrunsel et al., 2004). This process has been demonstrated experimentally, with moral disengagement serving as a mediator in the slippery slope effect (Welsh et al., 2015).

Ego Depletion. Baumeister (1998) proposed ego depletion theory. Human psychological energy is limited; when one repeatedly exerts self-control to resist instincts and habits over a period, substantial psychological energy is consumed, leading to temporarily decreased self-control capacity and resulting in **ego depletion** (Baumeister, Bratslavsky, Muraven, & Tice, 1998). Individuals in a

state of ego depletion have lower self-control, tend toward shallow cognitive processing, accept information consistent with existing concepts while ignoring inconsistent information, and are more likely to engage in unethical behavior (Fischer, Greitemeyer, & Frey, 2007). For example, sleep-deprived individuals have reduced self-control, leading to increased unethical behavior (Barnes, Gunia, & Wagner, 2015). Additionally, Kouchaki proposed the morning morality effect, suggesting people are more moral in the morning (Kouchaki & Smith, 2014), with more unethical behavior occurring in the afternoon and a more pronounced slippery slope effect (O'Connell & Kouchaki, 2014). The reason is that many things throughout the day consume our psychological energy, leading to gradually weakened self-control (Hofmann, Meindl, Mooijman, & Graham, 2018; Wilkowski et al., 2018). Psychological energy is depleted between repeated unethical behaviors, making ego depletion one of the driving factors of the slippery slope effect. Under continuous moral dilemmas, individuals' mental capacity is gradually consumed, leading to loss of reason, weakened self-control, and falling into the moral slippery slope effect.

Implicit Biases. Implicit biases refer to individuals being unaware of their own behavior, exhibiting an automatic state during moral judgment processes that produces unintentional unethical behavior (Kern & Chugh, 2009). Our moral decisions are influenced by systematic biases outside conscious awareness, which make our thinking and decision-making processes more comprehensive and holistic (Banaji et al., 2003). However, this also easily detaches our moral decisions from rationality, falling into unconscious, automatic decision-making states (Reynolds, 2006). This perspective suggests the slippery slope effect is facilitated by automated cognition resulting from implicit biases. Implicit biases make individuals unaware of unethical events before them, presenting an automatic judgment state that constitutes an internal cause of the slippery slope effect.

Tenbrunsel views the process of implicit biases as similar to the induction mechanism in mathematics: if the proposition holds for $n=1$ and $n=m$, it can be deduced that the proposition also holds for $n=m+1$. This process resembles the domino effect—proving that if any domino falls, the adjacent next domino will also fall, leading to the conclusion that “all dominoes will fall.” Tenbrunsel regards this induction mechanism as the internal process of the slippery slope effect, believing people use past practices as a baseline for evaluating new practices. If our past behavior had no moral problems, then similar new behaviors with only slight changes are likewise feasible. Implicit biases make it difficult for people to notice differences between behaviors, causing behavior to slip into irrational, automated states (Tenbrunsel et al., 2004).

Gino tested the role of implicit biases in the slippery slope effect. After participants engaged in unethical behavior, they were asked to complete a word-completion task. Due to implicit biases, people in the slippery slope condition had more difficulty activating thoughts related to moral standards. For example, when asked to complete the word “ch_____,” if they were aware they were

cheating, they would be more likely to fill in “cheat” rather than “chime.” Results showed that participants in the slippery slope condition used significantly fewer words related to deception and spent less time considering them than those in the abrupt condition (Gino et al., 2009), indicating that implicit biases make people less aware of the moral nature of behavioral differences, making it progressively easier to engage in unethical behavior.

3.2 Emotion: The Adaptation Perspective

In addition to loss of reason, people in the slippery slope effect typically also experience reduced emotional responses. To adapt to the psychological burden brought by cognitive and emotional dissonance, the self-regulation system continuously adjusts cognition or emotion to achieve dynamic balance. During this process, people experience changes in self-cognition and emotional numbness, states that can trigger the escalation of unethical behavior and produce the slippery slope effect. From the adaptation perspective, the psychological factors of the slippery slope effect include ethical numbing and self-dehumanization.

Ethical Numbing. Ethical numbing can be said to be a result of adaptation. On one hand, psychological research shows that repeated exposure leads to desensitization: after repeatedly experiencing sadness or suffering, people’s emotions become numb, such that those who have experienced much suffering are less sensitive to new pain. When repeatedly facing moral dilemmas, the dullness resulting from adaptation causes people to engage in potentially more unethical behavior—a process called **ethical numbing** (Bok, 1989). Ethical numbing makes it difficult for people in new situations to notice the ethical nature of behavior, leading to increasingly more unethical behavior. On the other hand, previous research shows that committing unethical behavior typically produces negative emotions such as frustration or guilt (Kayyal, Pochedly, McCarthy, & Russell, 2015). That is, lying is not easy, and the accompanying unbearable emotional experiences make people stop unethical behavior, make amends, or even self-punish (Regan, 1971; Yoel, Pizarro, Thomas, & Dan, 2013). Garrett (2016) suggested that people fall into the slippery slope effect possibly as a result of emotional adaptation. Functional magnetic resonance imaging (fMRI) scans during the slippery slope process found that participants’ amygdalae gradually became less active over time, and the amygdala is a crucial brain region for negative emotions. Researchers thus speculated that repeated lying leads to numbness in perceiving negative emotions, thereby producing the slippery slope effect (Garrett et al., 2016).

Self-Dehumanization. Moral theories related to human nature suggest that moral standards and moral emotions are unique human properties that distinguish us from animals (Haslam, 2006). Researchers typically describe the state of easily violating moral standards or severely lacking moral reason and emotion as “dehumanization,” initially used to describe brutal rulers during wartime. Recently, researchers have begun focusing on **self-dehumanization**, referring to a state in which individuals feel they lack humanity (Kouchaki, Dobson, Waytz,

& Kteily, 2018; Yang, Jin, He, Fan, & Zhu, 2015). For example, after participants intentionally attacked or excluded others in video games, they exhibited self-dehumanization, feeling less human (Bastian, Jetten, & Radke, 2012).

Kouchaki et al. revealed a bidirectional path between unethical behavior and self-dehumanization across eight experiments, finding that they are mutually causal. That is, after unethical behavior, people feel their humanity is less than normal, leading to self-dehumanization. Conversely, the result of self-dehumanization subsequently triggers further unethical behavior, creating a downward spiral (Kouchaki et al., 2018) that causes increasingly severe unethical behavior and produces the slippery slope effect. Therefore, to some extent, self-dehumanization helps explain the slippery slope effect of unethical behavior.

People who hold self-dehumanization beliefs (such as criminals) may continue down the path of immorality, making solutions that restore human emotions crucial. Eliminating self-dehumanization makes people believe they are complete individuals, motivating them to play social roles that conform to moral rules. Previous research shows that negative news in the moral domain can intensify people's unethical behavior (Yao & Yu, 2016). If media focus not on correcting wrongdoing by unethical individuals but on emphasizing this dehumanization (for example, comparing capitalists to vampires), it may cause these individuals to deepen their dehumanization beliefs and subsequently produce more unethical behavior.

4. Interventions for the Slippery Slope Effect

Targeting the external behavioral manifestation of the slippery slope effect—“small evils gradually escalating into major evils”—and its internal psychological causes of declining rational and emotional levels, research shows that effective interventions can be implemented from two aspects. First, activating prevention-focused motivation to control the source of unethical behavior and awaken subsequent rationality and emotions. Second, disseminating positive descriptive social norms, so that even when people temporarily fall into the irrational, unconscious, automated state of the slippery slope effect, they can promptly adjust their misconduct due to a good group moral atmosphere.

4.1 Prevention-Focused Motivation

In moral decision-making, people with different motivations act differently. Regulatory focus theory suggests that individuals with a **promotion focus** have a risk-seeking tendency and actively pursue goals, while those with a **prevention focus** have a vigilant tendency, striving to avoid failures and errors and focusing more on fulfilling personal responsibilities and obligations (Higgins, 1987). Prevention focus can reactivate self-regulation processes to prevent internal factors like moral disengagement from relaxing one's moral standards. Even when unethical behavior is trivial, prevention-focused individuals are more cautious in decision-making. This vigilance can help people avoid or promptly escape

irrational, unconscious, and automated states.

Research shows that activating prevention-focused motivation can reduce the slippery slope effect from its source. Gino found across four studies that prevention-focused individuals avoid risks, are more cautious when reporting performance, and engage in less unethical behavior (Gino et al., 2009). Manipulation of prevention focus typically begins before the formal experiment (Li et al., 2011; Welsh et al., 2015). First, the prevention focus group is presented with prevention-related words (e.g., safety, vigilance, prevention), while the control group receives neutral words. Next, prevention focus participants write about outcomes they most want to avoid and strategies to prevent those consequences, while control participants describe work or life scenarios. In the final minutes of the priming task, participants are asked to continuously feel fear and consider upcoming challenges. Results show that prevention focus participants engage in significantly less unethical behavior than controls and do not exhibit the slippery slope effect.

Bazerman argued that when unethical behavior develops gradually, the slippery slope effect dampens our awareness. Therefore, even trivial violations should be met with vigilance and immediate adjustment (Bazerman & Tenbrunsel, 2011). Enhancing prevention focus helps people maintain a morally sound status quo and is an effective method for controlling initial unethical behavior and detecting and adjusting subsequent behavior. Thus, even minor unethical behavior should be corrected and warned against on the spot. This can help people reactivate the reason and negative emotions gradually lost due to moral disengagement, restoring them to normal levels.

4.2 Descriptive Social Norms

Previous research on the slippery slope effect has focused primarily on individual behavioral change, but people do not live in a vacuum—their behavior is largely influenced by groups and society. While the slippery slope effect plunges people into irrational, emotionally numb, and automated states, it also amplifies group influence. The “conformity effect” that emerges at this moment is extremely critical and dangerous: good group moral norms can promptly guide people back on track, while the absence of good group moral norms can easily lead to large-scale “moral landslides.”

People’s behavior is influenced by social norm information. Currently, there are two types of norms: **descriptive norms** and **injunctive norms** (Cialdini, Reno, & Kallgren, 1990). Descriptive social norms convey information about what most people choose to do in a given situation, referring to “normal” behavior in that context; injunctive norms present prohibitive attitudes toward an event in current culture, reflecting a moral belief that “something should not be done.” Because descriptive social norms are easier to manipulate, they have become researchers’ preferred choice. Studies show that using positive descriptive social norms can effectively reduce unethical behavior. Köbis found in research on

official corruption that information about “the prevalence of bribery” greatly influenced participants’ subsequent decisions to engage in bribery. The researcher thus pointed out that conveying positive descriptive social norms is one of the most effective means of reducing corruption, suggesting that information about most people’ s choices can become a force influencing participants’ unethical behavior (Köbis et al., 2017). This demonstrates that disseminating positive, active descriptive social norms can effectively reduce the slippery slope effect and shape a good moral environment.

Today, media and the internet are the primary channels for disseminating descriptive social norms, while in-groups such as enterprises and schools are the most effective pathways. On one hand, increasing the promotion of descriptive social norms with good moral qualities and being cautious in reporting negative events helps improve citizens’ moral levels. On the other hand, organizations like enterprises or schools must consider whether they have deeply rooted moral cultures, whether unethical behavior has clear definition standards, and must eliminate or promptly correct unhealthy tendencies to establish high-quality descriptive moral norms. In short, because people tend to obey social norms, clear social moral norms should be established from the beginning, with explicit regulations and role models, so that people will maintain morality in the future.

5. Conclusion and Future Research Directions

Although recent explorations of the internal mechanisms and preventive measures of the slippery slope effect have yielded initial results, empirical research remains in a preliminary stage of development. Recently, some scholars have even questioned the existence of the slippery slope effect and proposed the **steep-cliff metaphor** (Kobis et al., 2017). They argue that people, motivated to maintain a good self-image, will not yield to small benefits to avoid the tension and stress brought by repeated unethical behavior (Teper, Inzlicht, & Page-Gould, 2011). Only extremely tempting benefits can overwhelm people’ s will (Francesca & Dan, 2012), like suddenly jumping off an extremely unethical steep cliff. This view breaks traditional notions, and future research could start from this contradiction to explore the respective applicable scopes and key boundary conditions of the slippery slope and steep-cliff metaphors. Additionally, there remains much room for exploration in research on the psychological and neural mechanisms of the slippery slope effect, with potential entry points described below.

5.1 Broadening Psychological Process Research on the Slippery Slope Effect

Overall, while some psychological mechanisms of the slippery slope effect have been confirmed, unexplored areas require continued investigation. Directions for inquiry include the subsequent trajectory of the slippery slope effect, empirical verification of internal causes, and its applicable scope. Future research should

focus on the temporal continuation effects of unethical behavior to deepen understanding of how unethical behavior changes over time. Specific approaches include:

First, investigating whether there are critical values or nodes in the slippery slope effect where people make completely different moral choices on either side. For example, Effron et al. found that time urgency affects the slippery slope process: near submission deadlines, participants' cheating rates rose sharply, while in unexpectedly extended time, cheating rates fell dramatically (Effron, Bryan, & Keith, 2015). Does the slippery slope effect thus depend on time urgency? Additionally, Nobel laureate Simon proposed the satisficing principle in decision-making: when pursuing benefits, people tend to follow a principle of satisfaction, stopping once they reach their expected gains rather than maximizing benefits (Simon, 1978). Since the slippery slope effect is mostly linked to monetary benefits, will it persist when monetary gains reach participants' satisfaction levels? Future research could explore critical points of the slippery slope effect from angles such as time properties and satisfaction, using relevant paradigms.

Second, current research on internal causes of the slippery slope effect has emotional adaptation remaining at the level of discussion and speculation. Subsequent research could indirectly examine emotional effects by investigating the degree of individual consciousness awakening, as emotional adaptation is typically accompanied by reduced consciousness awakening. Alternatively, combining implicit and explicit emotion measures, researchers could insert easy-to-understand, quick, and efficient emotion tests between task rounds for immediate detection. For example, implicit emotion measurement could use Huntsinger et al.'s (2009) information processing style test: if participants focus on the whole, they are in a neutral or positive emotional state; if they focus on parts, they are in a negative emotional state (Huntsinger, Lun, Sinclair, & Clore, 2009). Explicit emotion is typically self-reported, such as Russell's affect grid, which is sensitive to participants' emotional changes and does not cause fatigue with repeated use (Russell, Weiss, & Mendelsohn, 1989).

Third, will unethical behavior in one domain that shows a slippery slope pattern spread to other domains? For example, an employee who gradually becomes 消极怠工 (negative and lazy) might commit more work-unrelated unethical behaviors, such as taking office supplies home or spitting on the ground. Investigating whether unethical behavior in individuals' lives spreads to their work and study contexts with gradually intensifying trends is also an interesting topic. This could be examined by setting up two unrelated tasks in an experiment to observe whether unethical behavior in one task continues and escalates in another.

Fourth, while individual-level research has achieved certain results, interpersonal-level moral slippery slope research remains to be explored. Current interpersonal-level slippery slope research uses the two-party task and finds that people become increasingly likely to accept others' unethical behavior (Gino et al., 2009). However, this research has limitations: the participant

as a bystander is not the initiator of unethical behavior and does not bear responsibility for it. Future research should exclude the influence of diffusion of responsibility and make further improvements. Intergroup-level research remains at the theoretical level (Volkh, 2003) and has not yet been empirically tested. Given that interpersonal-level slippery slope effects involve immoral contagion (Tapp & Occhipinti, 2016) and that moral levels have a gradually decreasing tendency during transmission, future researchers could combine immoral contagion theories to improve and expand interpersonal-level, especially intergroup-level, slippery slope research.

5.2 Deepening Neural Mechanism Research on the Slippery Slope Effect

In recent years, researchers have begun investigating the neural mechanisms of the slippery slope effect. Garrett first used fMRI to examine these mechanisms, opening the neurological domain of slippery slope research (Engelmann et al., 2016). However, this study had too few participants and fMRI's temporal resolution was insufficient. Future research could combine fMRI with EEG (electroencephalography), which has stronger temporal resolution, along with participants' subjective reports, to reveal comprehensive psychological and physiological changes.

Following Garrett, researchers have continued exploring neural mechanisms of the slippery slope effect (Liu et al., 2019; Tang et al., 2019). For example, Liu et al. (2019) combined computational modeling with fMRI and used intranasal oxytocin to intervene in the slippery slope effect. Tang et al. (2019) used functional near-infrared spectroscopy (fNIRS) to detect the slippery slope effect, recording brain activity in the right dlPFC and rTPJ. Building on existing research, future neural mechanism studies could advance in the following ways.

First, slippery slope research could be extended to neural network interactions. Researchers have long proposed that unethical behavior should not be studied by focusing only on specific brain regions (Ganis, Kosslyn, Stose, Thompson, & Yurgelun-Todd, 2003). Both dorsolateral and dorsomedial prefrontal cortices are important nodes involved in evaluating moral severity (Baumgartner, Fischbacher, Feierabend, Lutz, & Fehr, 2009; Dogan et al., 2016). Research shows that signals from the amygdala to the broader moral-related network gradually weaken during adaptation to unethical situations, and this entire network may be related to the slippery slope effect (Bartra, McGuire, & Kable, 2013). Therefore, future research should examine whether the slippery slope effect is accompanied by changes in certain neural networks.

Second, serotonin could be used to investigate emotional adaptation in the slippery slope effect. Serotonin, also known as 5-hydroxytryptamine, is an inhibitory neurotransmitter believed to be related to emotional inhibition, reducing negative emotions generated by the amygdala in response to unethical behavior (Murphy, Norbury, Sullivan, Cowen, & Harmer, 2009). In the moral slippery

slope effect, researchers could observe whether serotonin secretion increases or temporarily promote serotonin transmission with medication to observe whether it enhances the slippery slope effect.

Third, because people wish to maintain a good moral self-image, deception typically causes physical and mental conflict (Eisenberg, 2000). In addition to the internal psychological causes summarized in this paper, this also causes certain physiological or behavioral changes in participants. Future research could use polygraphs or video recording in experiments to examine participants' physiological changes or variations in expressions and body posture to indirectly infer their psychological states.

Fourth, regulating cerebral cortical neurons could help suppress or interrupt the slippery slope effect. For example, transcranial direct current stimulation (tDCS), a non-invasive brain neuromodulation technology using electrical signals, could be used to intervene in the slippery slope effect. Although researchers have used tDCS to explore causal effects on the brain mechanisms of the slippery slope effect, the brain regions they stimulated were inconsistent, namely rTPJ (Noguchi & Oizumi, 2018) and dlPFC (Marechal, Cohn, Ugazio, & Ruff, 2017). Based on existing moral research, directly stimulating a single brain region cannot encompass all areas that may function under the slippery slope effect. For instance, previous research also shows that under tDCS, enhanced mPFC activity can similarly awaken participants' moral sense and emotional responses (Yuan et al., 2017). Therefore, future research could successively use fMRI and tDCS within the same study to systematically explore causal relationships between the slippery slope effect and certain brain regions.

References

Chinese References:

Chen, Y. F., & Mao, N. (2011). An experimental study on change blindness, slippery slope effect, and supplier ethics management. *Soft Science*, 25(9), 125-129.

Zheng, Q. Q. (2009). Research on corporate moral slippery slope behavior from a dynamic perspective. *Economic Theory and Business Management*, 11, 69-75.

English References:

Ashforth, B. E., & Anand, V. (2003). The normalization of corruption in organizations. *Research in Organizational Behavior*, 25(19), 1-52.

Barnes, C. M., Gunia, B. C., & Wagner, D. T. (2015). Sleep and moral awareness. *Journal of Sleep Research*, 24(2), 181-188.

Banaji, M. R., Bazerman, M. H., & Dolly, C. (2003). How ethical are you? *Harvard Business Review*, 81(12), 56-64.

- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Development*, 67(3), 1206-1222.
- Bartra, O., Mcguire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage*, 76(1), 412-427.
- Baron, R. A., Zhao, H., & Miao, Q. (2015). Personal motives, moral disengagement, and unethical decisions by entrepreneurs: Cognitive mechanisms on the “slippery slope” . *Journal of Business Ethics*, 128(1), 107-118.
- Bastian, B., Jetten, J., & Radke, H. R. M. (2012). Cyber-dehumanization: Violent video game play diminishes our humanity. *Journal of Experimental Social Psychology*, 48(2), 486-491.
- Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality & Social Psychology*, 74(5), 1252-1265.
- Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., & Fehr, E. (2009). The neural circuitry of a broken promise. *Neuron*, 64(5), 756-770.
- Bazerman, M. H., & Tenbrunsel, A. E. (2011). Ethical breakdowns. *Harvard Business Review*, 89(4), 58-65.
- Bok, S. (1989). *Secrets: On the ethics of concealment and revelation*. New York: Vintage.
- Burger, J. M. (2009). Replicating milgram: Would people still obey today? *American Psychologist*, 64(1), 1-11.
- Cain, D. M., Loewenstein, G., & Moore, D. A. (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *Journal of Legal Studies*, 34(1), 1-25.
- Chugh, D., Kern, M. C., Zhu, Z., & Lee, S. (2014). Withstanding moral disengagement: Attachment security as an ethical intervention. *Journal of Experimental Social Psychology*, 51(2), 88-93.
- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality & Social Psychology*, 58(6), 1015-1026.
- Courtois, C., & Gendron, Y. (2017). The “normalization” of deviance: A case study on the process underlying the adoption of deviant behavior. *Auditing: A Journal of Practice & Theory*, 36(3), 15-43.
- Detert, J. R., Trevino, L. K., & Sweitzer, V. L. (2008). Moral disengagement in ethical decision making: A study of antecedents and outcomes. *Journal of Applied Psychology*, 93(2), 374-391.

- Dogan, A., Morishima, Y., Heise, F., Tanner, C., Gibson, R., Wagner, A. F., & Tobler, P. N. (2016). Prefrontal connections express individual differences in intrinsic resistance to trading off honesty values against economic benefits. *Scientific Reports*, 6, 33263.
- Effron, D. A., Bryan, C. J., & M Keith, J. (2015). Cheating at the end to avoid regret. *Academy of Management Annual Meeting Proceedings*, 109(3), 395-414.
- Engelmann, J. B., & Fehr, E. (2016). The slippery slope of dishonesty. *Nature Neuroscience*, 19(12), 1543-1544.
- Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology*, 51, 665-697.
- Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—An experimental study on cheating. *Journal of the European Economic Association*, 11(3), 525-547.
- Fischer, P., Greitemeyer, T., & Frey, D. (2007). Ego depletion and positive illusions: Does the construction of positivity require regulatory resources? *Personality & Social Psychology Bulletin*, 33(9), 1306-1321.
- Francesca, G., & Dan, A. (2012). The dark side of creativity: Original thinkers can be more dishonest. *Journal of Personality & Social Psychology*, 102(3), 445-459.
- Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., & Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: An fMRI investigation. *Cerebral Cortex*, 13(8), 830-836.
- Garrett, N., Lazzaro, S. C., Ariely, D., & Sharot, T. (2016). The brain adapts to dishonesty. *Nature Neuroscience*, 19(12), 1727-1732.
- Gino, F., & Bazerman, M. H. (2009). When misconduct goes unnoticed: The acceptability of gradual erosion in others' unethical behavior. *Journal of Experimental Social Psychology*, 45(4), 708-719.
- Gino, F., Norton, M. I., & Ariely, D. (2010). The counterfeit self: The deceptive costs of faking it. *Psychological Science*, 21(5), 712-720.
- Gino, F., Schweitzer, M. E., Mead, N. L., & Dan, A. (2011). Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational Behavior & Human Decision Processes*, 115(2), 191-203.
- Henderson, A., Thoelen, G., Nadler, A., Barraza, J., & Nave, G. (2018). Testing the influence of testosterone administration on men's honesty in a large laboratory experiment. *Scientific Reports*, 8, 10.
- Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, 94(3), 319-340.
- Hippel, W. V., Lakin, J. L., & Shakarchi, R. J. (2005). Individual differences in motivated social cognition: The case of self-serving information processing.

Personality & Social Psychology Bulletin, 31(10), 1347–1357.

Hofmann, W., Meindl, P., Mooijman, M., & Graham, J. (2018). Morality and self-control: How they are intertwined and where they differ. *Current Directions in Psychological Science*, 27(4), 286–291.

Huntsinger, J. R., Lun, J., Sinclair, S., & Clore, G. L. (2009). Contagion without contact: Anticipatory mood matching in response to affiliative motivation. *Personality and Social Psychology Bulletin*, 35(7), 909–922.

Kayyal, M. H., Pochedly, J., McCarthy, A., & Russell, J. A. (2015). On the limits of the relation of disgust to judgments of immorality. *Frontiers in Psychology*, 6, 951.

Kern, M. C., & Chugh, D. (2009). Bounded ethicality: The perils of loss framing. *Psychological Science*, 20(3), 378–384.

Köbis, N. C., van Prooijen, J. W., Righetti, F., & Van Lange, P. A. (2015). “Who doesn’ t?” –The impact of descriptive norms on corruption. *Plos One*, 10(6), e0131830.

Kobis, N. C., van Prooijen, J. W., Righetti, F., & Van Lange, P. A. M. (2017). The road to bribery and corruption: Slippery slope or steep cliff? *Psychological Science*, 28(3), 297–306.

Kouchaki, M., & Smith, I. H. (2014). The morning morality effect: The influence of time of day on unethical behavior. *Psychological Science*, 25(1), 95–102.

Kouchaki, M., Dobson, K. S. H., Waytz, A., & Kteily, N. S. (2018). The link between self-dehumanization and immoral behavior. *Psychological Science*, 29(8), 1234–1246.

Lafollette, H. (2005). Living on a slippery slope. *Journal of Ethics*, 9(3/4), 475–499.

Li, A., Evans, J., Christian, M. S., Gilliland, S. W., Kausel, E. E., & Stein, J. H. (2011). The effects of managerial regulatory fit priming on reactions to explanations. *Organizational Behavior & Human Decision Processes*, 115(2), 268–282.

Liu, Y., Li, S., Lin, W., Li, W., Yan, X., Wang, X., ...& Ma, Y. (2019). Oxytocin modulates social value representations in the amygdala. *Nature Neuroscience*, 22(4), 633.

Marechal, M. A., Cohn, A., Ugazio, G., & Ruff, C. C. (2017). Increasing honesty in humans with noninvasive brain stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, 114(17), 4360–4364.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6), 633–644.

- Milgram, S. (1974). *Obedience to Authority: An Experimental View*. New York: Tavistock.
- Moore, C. (2008). Moral disengagement in processes of organizational corruption. *Journal of Business Ethics*, 80(1), 129-139.
- Moore, C., Detert, J. R., Trevino, L. K., Baker, V. L., & Mayer, D. M. (2012). Why employees do bad things: Moral disengagement and unethical organizational behavior. *Personnel Psychology*, 65(1), 1-48.
- Moscovici, S., & Zavalloni, M. (1969). The group as a polarizer of attitudes. *Journal of Personality and Social Psychology*, 12(2), 125-135.
- Murphy, S. E., Norbury, R., O' Sullivan, U., Cowen, P. J., & Harmer, C. J. (2009). Effect of a single dose of citalopram on amygdala response to emotional faces. *The British Journal of Psychiatry*, 194(6), 535-540.
- Noguchi, Y., & Oizumi, R. (2018). Electric stimulation of the right temporoparietal junction induces a task-specific effect in deceptive behaviors. *Neuroscience Research*, 128, 33-39.
- O' Connell, A., & Kouchaki, M. (2014). In the afternoon, the moral slope gets slipperier. *Harvard Business Review*, 92(5), 34-35.
- Regan, J. W. (1971). Guilt, perceived injustice, and altruistic behavior. *Journal of Personality & Social Psychology*, 18(1), 124-132.
- Rees, M. R., Tenbrunsel, A. E., & Bazerman, M. H. (2019). Bounded ethicality and ethical fading in negotiations: Understanding unintended unethical behavior. *Academy of Management Perspectives*, 33(1), 26-42.
- Reynolds, S. J. (2006). Moral awareness and ethical predispositions: Investigating the role of individual differences in the recognition of moral issues. *Journal of Applied Psychology*, 91(1), 233-243.
- Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, 57(3), 493-502.
- Schauer, F. (1985). Slippery slopes. *Harvard Law Review*, 99(2), 361-383.
- Shalvi, S., Handgraaf, M. J. J., & De Dreu, C. K. W. (2011). Ethical manoeuvring: Why people avoid both major and minor lies. *British Journal of Management*, 22, S16-S27.
- Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin*, 37(3), 330-349.
- Simon, H. A. (1978). Rationality as process and as product of thought. *American Economic Review*, 68(2), 1-16.

- Tang, H. H., Zhang, S., Jin, T., Wu, H. Y., Su, S., & Liu, C. (2019). Brain activation and adaptation of deception processing during dyadic face-to-face interaction. *Cortex*, 120, 326–339.
- Tapp, C., & Occhipinti, S. (2016). The essence of crime: Contagious transmission from those who have committed moral transgressions. *British Journal of Social Psychology*, 55(4), 756–772.
- Tenbrunsel, A. E., & Messick, D. M. (2004). Ethical fading: The role of self-deception in unethical behavior. *Social Justice Research*, 17(2), 223–236.
- Teper, R., Inzlicht, M., & Page-Gould, E. (2011). Are we more moral than we think? Exploring the role of affect in moral behavior and moral forecasting. *Psychological Science*, 22(4), 553–558.
- Volokh, E. (2003). The mechanisms of the slippery slope. *Social Science Electronic Publishing*, 116(4), 1026–1057.
- Welsh, D. T., Ordonez, L. D., Snyder, D. G., & Christian, M. S. (2015). The slippery slope: How small ethical transgressions pave the way for larger future transgressions. *Journal of Applied Psychology*, 100(1), 114–127.
- Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences of the United States of America*, 112(34), 10651–10656.
- Wilkowski, B. M., Ferguson, E. L., Williamson, L. Z., & Lappi, S. K. (2018). (How) Does initial self-control undermine later self-control in daily life? *Personality and Social Psychology Bulletin*, 44(9), 1315–1329.
- Yang, W. Q., Jin, S. H., He, S., Fan, Q., & Zhu, Y. J. (2015). The impact of power on humanity: Self-dehumanization in powerlessness. *Plos One*, 10(5), 19.
- Yao, Z., & Yu, R. (2016). The spreading of social energy: How exposure to positive and negative social news affects behavior. *Plos One*, 11(6), e0156062.
- Yin, L. J., & Weber, B. (2019). I lie, why don't you: Neural mechanisms of individual differences in self-serving lying. *Human Brain Mapping*, 40(4), 1101–1113.
- Yoel, I., Pizarro, D. A., Thomas, G., & Dan, A. (2013). Moral masochism: On the connection between guilt and self-punishment. *Emotion*, 13(1), 14–18.
- Yuan, H., Tabarak, S., Su, W., Liu, Y., Yu, J., & Lei, X. (2017). Transcranial direct current stimulation of the medial prefrontal cortex affects judgments of moral violations. *Frontiers in Psychology*, 8, 1812.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.