
18. Morality in attitudes and persuasion: metacognitive processes of thought validation

Pablo Briñol and Richard E. Petty

This chapter focuses on the mechanisms by which morality can exert its effects on attitudes and persuasion. We emphasize the importance of metacognitive processes in this domain because it has not received as much attention as processes of primary cognition. Metacognition involves secondary reflections on primary cognition, or thinking about one's own initial thinking. As will be described shortly, morality and morality-relevant variables (for example, honesty, purity, moral emotions) can influence persuasion by affecting what people think about their own thoughts and attitudes, and what they subsequently do with those mental contents. After a brief review of prior work on morality in the context of attitudes and persuasion, we describe how recent research on the metacognitive process of thought validation can be useful in understanding the role of morality in attitude change.

MORALIZATION OF ATTITUDES

Attitudes (people's general evaluations of objects, people, and issues; Petty et al., 2019) provide important input to behavior, and they have been considered historically as the most indispensable concept in social psychology (Allport, 1935; Briñol and Petty, 2012). Some attitudes are more consequential than others, though, in being "stronger," which refers to their durability and impact (Krosnick and Petty, 1995; Luttrell and Sawicki, 2020). Traditional indicators of attitude strength capable of predicting which attitudes are more consequential include features such as attitude accessibility, certainty, elaboration, extremity, importance, involvement, knowledge, and structural consistency (Petty and Krosnick, 1995). Beyond these traditional indicators of attitude strength, the extent to which the attitude is based in one's morals is perhaps the most studied new dimension of strength identified in the last two decades (Skitka et al., 2005). This strength feature has been linked to nearly all of the strength consequences of interest (for example, morally based attitudes are generally more resistant to change than attitudes based in practicality; see Skitka et al., 2021, for a review on moral attitudes), although, as we explain shortly, some important exceptions exist.

As an example of attitudes with a moral basis, consider two individuals who both have a positive attitude toward fair trade products. However, one might perceive that positivity to stem from non-moral considerations (for example, it will be a better-quality product), whereas the other might perceive the positivity to stem from moral considerations (for example, it is ethically right to pay workers fair wages). Indeed, people vary in how much they perceive a moral basis for their attitudes toward consumer products as well as toward social and political issues, and mounting evidence has shown that the more people perceive a moral (versus non-moral) basis for an attitude, the more they engage in attitude-consistent behavior (see Luttrell and Petty, 2021). As described throughout this chapter, recent research has empha-

sized the metacognitive nature of moralization, showing that the mere perception that one's thoughts and attitudes are based in morality (even as induced via random assignment) can often increase their impact (for example, enhancing attitude–behavior correspondence, resistance to change, stability, and beyond; Luttrell et al., 2016).

MORAL BASIS OF ATTITUDES AND PERSUASION

As an opening example of the relevance of morality in the domain of attitude change, consider recent work on matching persuasive messages content to attitude bases. Overall, the idea behind matching techniques is that persuasion works best when the message is congruent with (or matched to) some dimension of the recipient, such as the basis of the attitude (see Petty et al., forthcoming, and Teeny et al., 2021, for reviews of matching effects in persuasion). This matching approach was originally shown to be effective in the domain of changing attitudes with different functional bases (for example, using a self-image-oriented message to change an attitude based on social concerns; Snyder and DeBono, 1985). In another early matching study, Edwards (1990) showed that for an affectively based attitude, an affective message produced even more change than for a cognitively based attitude.

Now consider how matching can apply to the moral basis of attitudes. As noted, morally based attitudes are particularly strong, capable of predicting behavior, and difficult to change (Skitka et al., 2021). In a series of studies applying the matching notion to this strength variable, Luttrell et al. (2019) began by asking participants to rate the extent to which their attitudes on various topics (recycling, legalizing marijuana) were based on their morals. Then, participants were randomly assigned to receive a message making moral arguments (for example, “recycling releases toxic fumes that are harmful to pets”) or practical ones (for example, “recycling requires an increase in trucks that increase traffic congestion”). As predicted, the practical message became increasingly less effective in influencing attitudes as the moral basis of those attitudes increased, consistent with the typical resistance shown to persuasion as attitudes become stronger. However, the moral message became increasingly more effective as the moral basis of attitudes increased. In fact, the moral message had an easier time in persuading people with stronger (more moral) rather than weaker (more practical) attitudes (see Luttrell and Petty, 2021, for a conceptual replication). This work on matching messages to the moral basis of attitudes reveals that it is possible to influence strong attitudes with appropriate techniques (see Petty et al., 2023, for additional examples). It is especially effective to match messages to the particular dimensions of morality on which people's attitudes are based (for example, matching moral foundations message content to recipient values; Feinberg and Willer, 2013, 2015; see Luttrell, forthcoming, for a review of moral matching and persuasion).

Another recent example of tailoring messages to the moral bases of attitudes involves research on one- versus two-sided messages. One-sided messages are the most common form of persuasion and only present arguments on the advocated side. In contrast, two-sided messages also acknowledge some good points on the opposite side, in addition to presenting arguments for the advocated side (Hovland et al., 1953). Recently, Xu and Petty (2022, 2024) found that using a two-sided rather than the more typical one-sided message could be particularly effective for individuals whose attitudes were morally based or otherwise strong. Xu and Petty reasoned that those who cared deeply about their attitudes (that is, people with strong attitudes) would be especially appreciative of a message that recognized some validity

to their side. Then, in accord with the principle of reciprocity (Cialdini and Goldstein, 2004), these individuals would be more willing to acknowledge some validity to the advocated side. Although everyone would presumably show some appreciation for their side being respected, the acknowledgement coming from a two-sided message should be especially appreciated by those with strong attitudes.

In one study of this paradigm (Xu and Petty, 2022), the moral bases of people's attitudes toward gun control were assessed, and then they were presented with a persuasive message that advocated against their own view. The message was either entirely one-sided or presented both sides. That is, the two-sided message presented the same arguments as in the one-sided message, but then also presented some arguments on the recipient's side, before concluding that the side advocated for was superior. To examine the reciprocity hypothesis, the recipient's openness (receptiveness) to the side advocated was assessed (see Briñol and Petty, 2023; Hussein and Tormala, 2021; Minson and Chen, 2022, for recent reviews on openness and persuasion). The results showed that for the one-sided message, the typical attitude strength effect occurred: people were more resistant to the message the more morally based their attitudes were. However, for the two-sided message, the opposite occurred: the more morally based (stronger) the attitude was, the more open it was to the other side.

In this study, Xu and Petty (2022) also showed that appreciation for the author mediated this openness effect, and furthermore, the more open to the other side the recipient reported being, the more favorable their attitudes became toward the advocated position. This interaction between moral basis and attitude strength on openness was also shown for a message advocating mask wearing during Covid-19 for people who were initially against it. In sum, although it is generally the case that strong attitudes are more resistant to change than weak ones, this recent work on morality and persuasion has demonstrated that there are particular techniques that can be effective in rendering strong attitudes more open to change.

PERCEIVED MORALITY AS A METACOGNITIVE ASSESSMENT

In the studies just described, participants were asked to report the moral basis of their attitudes, though as will be described next, similar outcomes are obtained when people are randomly assigned to believe that their attitudes are based in morality (Luttrell et al., 2016). Believing that the origin of an attitude is morally based or not is a metacognitive assessment because it involves thinking about thoughts. That is, reflecting on the perceived origin of one's own mental constructs (including the basis of attitudes) involves secondary cognition (e.g., Gascó et al., 2018; Goupil and Kouider, 2019). As will be described throughout, these secondary, metacognitive assessments about attitudes and other mental constructs are consequential in determining the extent to which initial thoughts are used to guide further thinking, judgment, and action (Petty et al., 2007).

Much persuasion work on metacognitive processes has been guided by Self-Validation Theory (SVT) (Briñol and Petty, 2022). Therefore, we rely on SVT in organizing the subsequent sections of this chapter. SVT deals with what people do with their thoughts, and especially whether their thoughts become consequential. The key notion of SVT is that the greater the perceived validity of one's thoughts, the more they are translated into overall judgments; and the greater the perceived validity of one's judgments, the more likely they are to guide behavior (Petty et al., 2002). As described next, SVT focuses on a subjective sense that one's

thoughts are appropriate (valid) to use, and this assessment is influenced by thought-relevant as well as incidental factors (for example, one's emotions) completely unrelated to the content of thoughts.

PERCEIVED MORALITY IS MALLEABLE AND VALIDATING

Perceptions of morality are malleable, and these perceptions can serve as a validating variable of thoughts and attitudes. For example, Luttrell et al. (2016) showed that perceptions of morality can vary depending on the person and the situation, and that merely perceiving a moral (versus non-moral) basis to one's attitudes can increase attitude strength, as indicated by greater attitude to behavior correspondence and greater resistance to change. In one study by Luttrell et al. (2016), participants were first asked to list the thoughts that they had about recycling. Next, participants were provided with bogus feedback regarding the basis of those thoughts about recycling, being told that their thoughts related either to morality or to a control value. Specifically, participants were told that the thoughts they listed were submitted to a computerized analysis that compared them to thoughts generated by other students (a manipulation adapted from prior SVT research; Petty et al., 2002). After an eight-second delay, introduced to make the cover story more plausible, participants received the feedback information. Those in the moral basis condition were told that their thoughts were especially based on morality, whereas those in the non-moral basis condition were told that their thoughts were especially based on other control values, such as tradition or practicality (control values were varied depending on the study). Upon receiving this feedback, participants reported their attitudes toward the topic for which they initially listed their thoughts (recycling) along with a collection of attitude strength variables, including relevant behavioral intentions toward recycling.

As predicted, the results obtained by Luttrell et al. (2016) demonstrated that leading people to view their attitudes as based in morality rendered the attitude more consequential, as indicated by stronger correspondence between the attitudes and relevant behavior and less attitude change following a persuasive message. Subsequent research using this paradigm in which perceived morality is manipulated experimentally has also shown that the mere perception that one's attitude is based in morality can increase the attitude's strength in different applied settings—for example, enhancing attitude–behavior correspondence toward consumer products (Luttrell and Togans, 2021), and increasing the consistency of thoughts with attitudes toward legalizing several doping behaviors in sports (Horcajo and Gil, 2024)—with additional consequences for other attitude strength outcomes (for example, attitude stability; Luttrell and Togans, 2021).

COGNITIVE VALIDATION

Perceiving one's thoughts and/or attitudes as morally based can serve as a validating variable, as demonstrated in the examples provided. That is, perceived morality can validate thoughts and attitudes, even if those initial thoughts (for example, about recycling, gun control, and so on) are not necessarily related to morality by default. As the studies described demonstrated, moralizing mental contents increased their subsequent impact on judgment and behavior. In

the next examples, morality and morality-relevant variables play a different role. This section focuses on cases in which mental contents potentially relevant to morality serve as a primary cognition to then be validated by other variables. As will be described shortly, morality can be a validating variable when it takes the role of secondary cognition, but it can also be relevant to the initial thoughts people generate that are then validated.

First consider SVT work on how validating socially relevant thoughts impact moral decisions (DeMarree et al., 2012). In one experiment, participants were first primed with the goal to cooperate (high in morality) or compete (lower in morality). This priming induction served as the primary cognition. Next, participants were instructed to reflect on previous times when they held high or low levels of power. This recall task was the validating induction of the initial primes because power was introduced after the priming task and because power is often associated with confidence and pleasantness (see Briñol et al., 2017, for a review on power). That is, power served as the secondary cognition designed to validate initial primes. Following these two inductions, the impact of the initial primes on people's behavior in two economic decision-making tasks was assessed: that is, the dictator game (Bolton et al., 1998) and the trust game (Berg et al., 1995).

Consistent with the idea that validity processes can influence goal pursuit, there was more goal-congruent behavior among individuals made to feel powerful rather than powerless. For those in the powerless condition, the impact of the primes on behavior was invalidated to the point that it was eliminated (see Briñol and Petty, forthcoming, for a review on when invalidation eliminates, attenuates, or reverses the impact of mental contents). In subsequent studies, the same validation processes have been applied to various kinds of primary cognitions that can be considered to be low morality, such as thoughts about hurting others (DeMarree et al., 2014; Santos et al., 2019, 2022), thoughts about discriminating against people with disabilities (Requero et al., 2020), and prejudiced thoughts against refugees (Moreno et al., 2021).

Next, consider work on thought validation by priming morality. Across two experiments by Santos and Rivera (2015), participants first were paired with either a prosocial or antisocial target. In the pro-social condition, participants read an interview presenting a morally good neighbor who collected money to help other people who could not pay their rent. In contrast, in the antisocial version of the interview, the neighbor stole the money from a lost wallet that he found in the lift. This induction of thought-valance served as the primary cognition to be validated. Next, participants received either a morally relevant prime (a prime about justice) or a control prime. Specifically, prime words were associated with ethics and morality (for example, honesty, integrity, righteous, loyal, truthful), or with legal terms that served to prime the concept of justice (for example, norm, human rights, judicial system, law, democracy, public order), or with neutral words that served as controls (for example, box, detergent, computer). This priming induction was the validating variable. Finally, participants reported their intentions to punish the target of the initial description by taking money from him in an economic game (Experiment 1) or by assessing their behavioral intentions to punish the target (Experiment 2). Results indicated that punishment assignment was greater for those participants who were paired with the antisocial target and had ethical concepts primed, than for those who had neutral concepts primed. Priming justice also increased punishment intentions relative to controls. In this research, the effects of ethics and justice did not differ from each other.

COGNITIVE AND AFFECTIVE VALIDATION

According to SVT, people can decide that their thoughts are valid to use for two general kinds of reasons: people can rely on their thoughts because they believe their thoughts are correct (cognitive validation), or because they feel good about them (affective validation). As described in the previous section, people make inferences of correctness when they are feeling certain about the validity of their thoughts, such as when the thoughts are inspired by priming moral concepts (Santos and Rivera, 2015), and by merely recalling past episodes of confidence (Requero et al., 2020). In addition to being certain about thoughts, feeling good about thoughts can also increase thought usage (affective validation). For example, people rely on their positive and negative thoughts more when they are feeling happy (versus sad, and versus neutral) following thinking (Briñol et al., 2007). If thoughts were positive, making people feel especially good would increase persuasion compared to a sad state. However, feeling good can also magnify the impact of negative thoughts by validating them, thereby decreasing persuasion (see Huntsinger, 2013; Paredes et al., 2013, for additional examples of affective validation).

Following initial work on validation through happiness versus sadness, recent research has found similar results for more moral emotions such as pride and shame. Specifically, Lamprinakos et al. (2022) showed that participants induced to feel pride following thinking used their previously generated positive and negative thoughts more in forming their attitudes, compared to those led to feel shame. Importantly, since emotions such as happiness and pride are associated with both pleasantness and certainty (and sadness and shame are both unpleasant and doubtful), it is not clear whether affective or cognitive validation was responsible for the validation effect in these studies.

In order to demonstrate that validation can occur via both cognitive and affective routes, we recently examined emotions that have the potential to affect certainty and pleasantness appraisals differently. For example, anger and disgust are unpleasant emotions that are associated with certainty, whereas surprise and awe are more pleasant emotions that are associated with doubt and uncertainty. Consistent with our differential appraisals hypothesis (Briñol et al., 2018), we demonstrated that each of these emotions is capable of inducing either more or less thought use depending on which appraisal (certainty or pleasantness) is dominant at the time. This approach concurs with previous frameworks in highlighting the importance of appraisals (Lerner and Keltner, 2000), but introduces some important novelties such as predicting that different appraisals can be relevant when varied even within the same emotion. Therefore, appraisals matter not only when comparing the effect of different emotions, but also when examining the impact of the very same emotion for which one appraisal or another is made salient.

In a recent illustration of the utility of the SVT differential appraisals hypothesis relevant to promoting egalitarian attitudes, participants were asked to think about the positive or negative aspects of a proposal to hire people with disabilities in an organization (Requero et al., 2020). Following this thought valence manipulation, participants were assigned to write about personal episodes in which they felt either hope or hopelessness. Hope and hopelessness were chosen because for these emotions the confidence and pleasantness appraisals are mismatched. Hope is a pleasant state that is associated with uncertainty, whereas hopelessness is unpleasant though associated with confidence. After participants generated their thoughts and received the emotion induction, they were exposed to an appraisal manipulation designed to focus them on a specific appraisal within the emotion (pleasantness or confidence) (Briñol et al., 2018).

Specifically, participants were required in this study to respond to questions that contained words related to pleasantness (pleasantness appraisal) or questions that contained words related to confidence (confidence appraisal). Finally, after the three manipulations (positive versus negative initial thoughts induction; hope versus hopelessness induction; pleasantness versus confidence appraisal induction), participants reported their attitudes toward the proposal to hire people with disabilities.

In line with SVT predictions, this research showed that feeling hopelessness or hope following thought generation can lead to different (and opposite) effects on the use of thoughts. This effect depends on whether the confidence or the pleasantness appraisal of emotions was made salient. When individuals were focused on the confidence/doubt appraisal of the emotion, then feeling hopeless led to more thought use than hope because experiencing hopelessness is associated with an appraisal of confidence that was misattributed to feeling sure about the accuracy or correctness of one's thoughts relative to feeling hope, an emotion associated with uncertainty (cognitive validation). In contrast, when individuals were focused on the pleasantness/unpleasantness appraisal of the emotion, then experiencing hope led to more thought use than feeling hopeless because experiencing hope is associated with an appraisal of pleasantness that was misattributed to feeling good about or liking one's thoughts relative to hopelessness, an emotion associated with feeling unpleasant (affective validation).

Beyond hope and hopelessness, recent research has applied this SVT differential appraisal notion to examine the impact of moral emotions for which appraisals are dissociated, such as admiration (pleasant but doubtful emotion) and disdain (unpleasant but confident emotion) (Lamprinakos et al., 2024a). When participants were induced to feel admiration in this series of studies, focusing on the confidence/doubt appraisal of that emotion reduced thought usage compared to focusing on the pleasantness/unpleasantness appraisal because admiration is associated with uncertainty. In contrast, when participants were induced to feel disdain and were focused on the confidence/doubt appraisal of that emotion, then experiencing disdain led to more thought use than focusing on the pleasantness/unpleasantness appraisal of that emotion because disdain is associated with confidence.

MORALITY AS AN OUTCOME OF VALIDATION PROCESSES

An important aspect of SVT is that it can apply to any accessible mental contents. That is, perceptions of validity apply to whatever the available mental contents are at the time, including all kinds of thoughts. In the examples described in the previous sections, we focused on thoughts relevant to promoting diversity and morally relevant choices. In this section, we describe additional SVT research demonstrating that the impact of any mental content (for example, group identity) can be magnified when validated, but attenuated when invalidated, affecting other moral decisions (for example, as assessed with common moral assessments such as the trolley dilemma).

In line with the idea that any mental content can be validated, recent research has shown that metacognitive confidence can improve the predictive ability of measures of group identity on moral decisions. Specifically, this research focused on the role of confidence in identity fusion. Identity fusion is a powerful feeling of connectedness to one's group that is capable of predicting extreme pro-group behavior (Swann et al., 2014). In relevant research, Paredes et al. (2020) demonstrated that assessing confidence in people's identity fusion responses can improve the

ability of the fusion measure to predict extreme, even immoral, pro-group outcomes. Across three studies, participants reported their level of identity fusion with their country, and their confidence in responses to the scale (predictor variables). Simulated sacrifice behavior (that is, responses to a moral trolley dilemma) and reported willingness to fight and die for one's country were the dependent measures. Specifically, participants were introduced to the "summoning the death train scenario" (Swann et al., 2014). This scenario informs participants that a runaway train was about to crush and kill five citizens of their country unless they flipped a switch that diverted the train to their own railway track, killing them, but leaving the five in-group members unharmed. Consistent with SVT, the results of this research revealed that confidence in fusion responses was capable of moderating the effects of identity fusion on all measures of sacrifice, with greater consistency between fusion and endorsement of pro-group self-sacrifice obtained for those with high (versus low) confidence.

Subsequent research has also shown how other situations that are indirectly related to confidence (high validity) or doubt (low validity) can affect the association between identity and subsequent moral behavior (e.g., Paredes et al., 2021). Taken together, this SVT research reveals that any mental content becomes more consequential for judgment and action as the perceived validity of the thoughts is increased, including the responses to standardized individual difference inventories (e.g., Shoots-Reinhard et al., 2015; Wichman et al., 2010). Thus, taking into consideration metacognitive confidence can be useful for understanding and predicting the effects of any mental content, related or unrelated to morality, including inventories assessing other forms of pro-social (e.g., Moreno et al., 2021) and antisocial behavior (e.g., Santos et al., 2019).

THE CASE OF CHEATING

The previous sections focused on how confidence can improve the ability of mental contents assessed with standardized instruments to guide moral decisions. In this section, we move to three other cases in which confidence can improve the ability of other mental constructs to predict a different set of morally relevant outcomes. As described next, thoughts about cheating and honesty can serve as primary cognitions to then be validated, and decisions to cheat can also be affected by validation of other relevant thoughts.

First, consider how popular media often present cases of people in high-power positions as involved in all kinds of immoral behaviors, such as cheating, breaking the law, and violating Covid-19 mandatory policies during the pandemic. Consistent with this anecdotal evidence, research findings support the association between feeling powerful and making immoral decisions. For example, Lammers et al. (2010) showed that participants induced to feel high (versus low) power cheated more, as indicated by lying about the number obtained after the roll of dice, to increase their chances to win a lottery in an experimental study.

Building on this link between power and dishonest behavior, Lamprinakos et al. (2024b) showed that power can not only increase but also decrease cheating, depending on the thoughts validated by the feelings of power. Participants in these experiments were first asked to generate either positive or negative thoughts about cheating. Following this initial manipulation of thought direction, participants were placed in either high- or low-power conditions by asking them to recall past episodes of high or low power (Galinsky et al., 2003). After the two inductions, cheating was measured using different paradigms depending on the study, such

as assessing infidelity intentions in their relationships, and telling lies to gain more money. Relative to powerless participants, those induced to feel powerful showed more reliance on the initial thoughts induced. Consequently, the effect of the direction of the thoughts on cheating was greater for participants with high (versus low) perceived power. Specifically, high power increased cheating only when initial thoughts were induced to be favorable toward cheating. That finding replicates the established effect observed in prior literature, but arguably through a different process. In contrast, power decreased cheating when it validated unfavorable thoughts about cheating, thus reversing the traditional effect.

In this initial research by Lamprinakos et al. (2024b), power served as the secondary cognition validating favorable or unfavorable thoughts about cheating, because power was introduced after participants had already generated their primary cognitions. However, power can also serve as the primary cognition when introduced in the first place, and people can vary in the confidence that they associate with their feelings of power. Consistent with this possibility, recent research by Toader et al. (2024) tested the extent to which the perceived validity associated with power could strengthen the ability of perceived power to guide immoral decisions.

In this research (Toader et al., 2024), participants were first asked to respond to the Generalized Sense of Power Scale (Anderson et al., 2012). The responses to the scale served as the primary cognition. After completing the power scale, participants were then asked to report the perceived validity of their responses to the scale. Perceived validity was assessed by asking participants to rate how confident they were in their responses to the initial power inventory. A second study of this series moved to an experimental design where power was manipulated at the beginning of the study by asking participants to recall a time in which they felt either powerful or powerless. This power manipulation varied momentary feelings of power (Galinsky et al., 2003). Then, perceived validity was manipulated by asking participants to recall past experiences of confidence versus doubt (Petty et al., 2002). As noted, this well-established manipulation can lead to momentary feelings of confidence, and those feelings can be misattributed to the mental content currently available in mind (feelings of power in this case). The dependent measure used in both studies was a “sneaking into the queue” paradigm (Paredes et al., 2021). Basically, participants during the Covid-19 pandemic were offered the possibility of being included in a priority vaccination list even if they were not eligible for the vaccine at that time (immoral behavior), or to decline that possibility and wait for their turn (moral behavior).

The results of these studies revealed that the perceived validity of power moderated the extent to which felt power predicted cheating versus honest responses. Consistent with SVT, participants with low perceived power (either measured or manipulated) exhibited less cheating behavior when perceived validity of their perceived power (either measured or manipulated) was relatively high rather than low. High (versus low) perceived validity increased cheating behavior when participants felt powerful compared to feeling powerless. Taken together, both studies are consistent with the idea that power corrupts, and feeling sure of felt power corrupts even more. That is, high perceived validity improved the ability to predict the impact of power on cheating. A third study on this series replicated and extended the results beyond the pandemic context and used actual cheating behavior.

As a final example of the potential of SVT to improve our understanding of immoral behaviors, consider a set of studies focused on how confidence can improve the predictive validity of other mental constructs capable of predicting cheating. This research was conducted by Santos et al. (2024), and built on prior work showing that some personality

traits such as Honesty-Humility and the Dark Triad are associated with cheating behavior. Honesty-Humility (H-H) is a personality trait that captures a person's sincerity, fairness, greed avoidance, modesty, and lower likelihood of exploitative or deceptive behavior (Ashton et al., 2014). As a result, this dimension has been associated with ethical behavior (Kleinlogel et al., 2018), and low antisocial behaviors such as cheating (Pfattheicher et al., 2019). In contrast, the Dark Triad is a cluster of three personality traits known for predicting antisocial and unethical tendencies in people characterized by a lack of empathy and a tendency to exploit or manipulate others for personal gain (Paulhus and Williams, 2002). Specifically, the Dark Triad consists of narcissism, Machiavellianism, and psychopathy.

The SVT research by Santos et al. (2024) was designed to build convergent validity about both the association between H-H and cheating behavior, as well as the association between the Dark Triad and cheating behavior. More uniquely, this research examined whether those associations between traits and immoral behavior could be strengthened or weakened by taking certainty in scale responses into consideration. Across three studies, participants began by completing the Honesty-Humility and Dark Triad scales, and then rated the certainty they had in their responses to each of those two inventories (predictor variables). As the dependent variables, Santos and colleagues used cheating behavior in a moral dilemma (study 1: Stag Hunt dilemma), a die-rolling paradigm (study 2), and a paradigm asking participants to rate the likelihood of cheating by providing random responses when serving as participants in different surveys (Experiment 3: measures adopted from Suri et al., 2011). As predicted, low scores in Honesty-Humility and high scores in the Dark Triad were associated with more cheating behavior across studies and measures, replicating previous research. More relevantly, it was predicted and found that certainty in responses moderated these main effects. Specifically, the link between each of these two traits and cheating was greater for participants with higher levels of certainty in their responses to the inventories. Importantly, these findings were replicated in a third sample where the two scales were administered to all participants and certainty was manipulated (rather than measured). In sum, factoring in the certainty with which individuals hold their traits can contribute to enhancing the predictive validity of these scales, and improving the ability to predict different kinds of unethical behaviors.

THE CASE OF PURITY

An important aspect of SVT is that it proposes that the meaning of variables which affect thought validity is flexible. Changing the meaning of any experience changes the impact of that experience on thought usage. Put simply, meaning will moderate the impact of the validating variable on the outcome of interest. Changing the meaning can change the outcome (see Briñol et al., 2018, for an SVT review on meaning).

As an illustration of how changing the default meaning of a variable can change its impact, consider the case of purity and cleansing. The meaning of cleansing actions (for example, washing one's hands, using cleansing products) is often associated with removing dirt. Consistent with this meaning, research has shown that cleaning actions following thinking were initially found to reduce (remove) the impact of negative thoughts, such as the guilt associated with a previous transgression and the effects of perceived threats (e.g., Lee and Schwarz, 2011; Zhong and Liljenquist, 2006). Because SVT views cleansing as a general variable associated with disliking something, it can be applied to positive and negative thoughts

alike. Subsequent research on cleansing confirmed this prediction (e.g., Florack et al., 2014; see Lee and Schwarz, 2020; Lee et al., 2024, for reviews).

More relevant to the meaning idea is that one would expect that if the default meaning of the action of cleansing was reversed, from removal of something bad (dirt) to adding something good (purity), the outcome would also be reversed. In an illustrative study, Kim et al. (2018) had participants think about a recent time when they did something wrong, and then gave them the opportunity to wash their hands. When the action of washing was framed as removing dirt (the default meaning), the results showed that guilt over the wrong action decreased, replicating the original effect of hand washing. In contrast, when the same action was framed as adding purification to the body, the experienced guilt increased, reversing the original effect. This outcome is consistent with the SVT assumption that it is the meaning of an action, and not the action itself, that matters.

In sum, the work presented in this section highlights how the meaning of the action relevant to purity can matter for validation processes. SVT also points to other moderators such as the timing of the inductions (for example, does cleansing precede or come after thought generation?), and the amount of thinking (for example, is the motivation and ability to think about the validity of thoughts high or low?) that are relevant for understanding the effect of purity and other morally relevant variables. In particular, SVT effects are more likely when the validating variable comes after rather than before the thoughts to be validated are generated, and when thinking is relatively high rather than low (see Briñol and Petty, 2022). Validation processes require more thinking than the operation of primary cognition processes alone, because not only must people have some thoughts to validate, but they also need some motivation and ability to care about the correctness or pleasantness of their previously generated thoughts.

SUMMARY

In order to anticipate when and how morality will increase or decrease attitude strength and persuasion, it is necessary to consider the processes by which variables associated with morality operate. After briefly giving some examples of the role of primary cognition in morality effects, this chapter focused specifically on metacognitive processes of validation and invalidation. Although morality-relevant variables (for example, purity, honesty) can serve as primary thoughts to be validated by variables associated with confidence (for example, power) and pleasantness (for example, pride), we also highlighted how these morally relevant variables can serve to validate or invalidate other kinds of previously generated thoughts. As noted, the initial thoughts that were validated in the studies of this review can relate to morality (for example, when they refer to cheating, transgressions, and so on) but they can also be totally unrelated to morality. Furthermore, the outcome of these validation processes can be both relevant to morality (for example, changing decisions in moral dilemmas) but also be unrelated to morality (for example, changing attitudes toward utilitarian products). Across all kinds of outcomes, primary cognitions, and through a wide array of validating variables, SVT offers a conceptual framework capable of shedding new light on the role of morality in the domain of attitudes and persuasion.

REFERENCES

- Allport, G.W. (1935). Attitudes. In C. Murchison (ed.), *Handbook of Social Psychology* (pp. 798–884). Clark University Press.
- Anderson, C., John, O.P., and Keltner, D. (2012). The personal sense of power. *Journal of Personality*, 80, 313–344. DOI: 10.1111/j.1467-6494.2011.00734.x.
- Ashton, M.C., Lee, K., and de Vries, R.E. (2014). The HEXACO honesty-humility, agreeableness, and emotionality factors: A review of research and theory. *Personality and Social Psychology Review*, 18(2), 139–152.
- Berg, J., Dickhaut, J. and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1), 122–142. DOI: 10.1006/game.1995.1027.
- Bolton, G.E., Katok, E., and Zwick, R. (1998). Dictator game giving: Rules of fairness versus acts of kindness. *International Journal of Game Theory*, 27, 269–299.
- Briñol, P., and Petty, R.E. (2012). The history of attitudes and persuasion research. In A. Kruglanski and W. Stroebe (eds), *Handbook of the history of social psychology* (pp. 285–320). Psychology Press.
- Briñol, P., and Petty, R.E. (2022). Self-validation theory: An integrative framework for understanding when thoughts become consequential. *Psychological Review*, 129(2), 340–367. DOI: 10.1037/rev0000340.
- Briñol, P., and Petty, R.E. (2023). Openness and persuasion: Multiple processes, meanings, and outcomes. In V.C. Ottati and C. Stern (eds), *Divided: Open-mindedness and dogmatism in a polarized world* (pp. 59–77). Oxford University Press.
- Briñol, P., and Petty, R.E. (forthcoming). Multiple effects of discarding thoughts through invalidation: Attenuation, elimination, reversals, and compensation. In P.J. Carroll, K. Rios, and K.C. Oleson (eds), *Handbook of the uncertain self* (2nd edn). Psychology Press.
- Briñol, P., Petty, R.E., and Barden, J. (2007). Happiness versus sadness as determinants of thought confidence in persuasion: A self-validation analysis. *Journal of Personality and Social Psychology*, 93, 711–727. DOI: 10.1037/0022-3514.93.5.711.
- Briñol, P., Petty, R.E., Durso, G.R.O., and Rucker, D.D. (2017). Power and persuasion: Processes by which perceived power can influence evaluative judgments. *Review of General Psychology*, 21, 223–241.
- Briñol, P., Petty, R.E., Stavradi, M., Lamprinakos, G., Wagner, B.C., and Díaz, D. (2018). Affective and cognitive validation of thoughts: An appraisal perspective on anger, disgust, surprise, and awe. *Journal of Personality and Social Psychology*, 114(5), 693–718. DOI: 10.1037/pspa0000118.
- Cialdini, R.B., and Goldstein, N.J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55, 591–621. DOI: 10.1146/annurev.psych.55.090902.142015.
- DeMarree, K.G., Briñol, P., and Petty, R.E. (2014). The effects of power on prosocial outcomes: A self-validation analysis. *Journal of Economic Psychology*, 41, 20–30.
- DeMarree, K.G., Loersch, C., Briñol, P., Petty, R.E., Payne, B.K., and Rucker, D.D. (2012). From primed construct to motivated behavior: Validation processes in goal pursuit. *Personality and Social Psychology Bulletin*, 38, 1659–1670. DOI: 10.1177/0146167212458328.
- Edwards, K. (1990). The interplay of affect and cognition in attitude formation and change. *Journal of Personality and Social Psychology*, 59, 202–216. DOI: 10.1037/0022-3514.59.2.202.
- Feinberg, M., and Willer, R. (2013). The moral roots of environmental attitudes. *Psychological Science*, 24(1), 56–62. DOI: 10.1177/0956797612449177.
- Feinberg, M., and Willer, R. (2015). From gulf to bridge: When do moral arguments facilitate political influence? *Personality and Social Psychology Bulletin*, 41(12), 1665–1681. DOI: 10.1177/0146167215607842.
- Florack, A., Kleber, J., Busch, R., and Stöhr, D. (2014). Detaching the ties of ownership: The effects of hand washing on the exchange of endowed products. *Journal of Consumer Psychology*, 24(2), 284–289. DOI:10.1016/j.jcps.2013.09.010.
- Galinsky, A.D., Gruenfeld, D.H., and Magee, J.C. (2003). From power to action. *Journal of Personality and Social Psychology*, 85, 453–466. DOI: 10.1037/0022-3514.85.3.453.
- Gascó, M., Briñol, P., Santos, D., Petty, R.E., and Horcajo, J. (2018). Where did this thought come from? A self-validation analysis of the perceived origin of thoughts. *Personality and Social Psychology Bulletin*, 44(11), 1615–1628. DOI: 10.1177/0146167218775696.

- Goupil, L., and Kouider, S. (2019). Developing a reflective mind: From core meta-cognition to explicit self-reflection. *Current Directions in Psychological Science*, 28(4), 403–408. DOI: 10.1177/0963721419848672.
- Horcajo, J., and Gil, R. (2024). The meta-cognitive role of thought morality in attitude change and behavioral intentions related to doping. *Journal of Sports Sciences*, 41(21), 1896–1905. DOI: 10.1080/02640414.2024.2307757.
- Hovland, C.I., Janis, I.L., and Kelley, H.H. (1953). *Communication and persuasion: Psychological studies of opinion change*. Yale University Press.
- Huntsinger, J.R. (2013). Incidental experiences of affective coherence and incoherence influence persuasion. *Personality and Social Psychology Bulletin*, 39(6), 792–802. DOI: 10.1177/0146167213482588.
- Hussein, M.A., and Tormala, Z.L. (2021). Undermining your case to enhance your impact: A framework for understanding the effects of acts of receptiveness in persuasion. *Personality and Social Psychology Review*, 25(3), 229–250. DOI: 10.1177/10888683211001269.
- Kim, T.W., Lee, S., Duhachek, A., Briñol, P., and Petty, R.E. (2018, January). The meaning of cleansing moderates the impact of cleansing on guilt reduction. Presented at the annual meeting of the Society Consumer Psychology, Sidney, Australia.
- Kleinogel, E.P., Dietz, J., and Antonakis, J. (2018). Lucky, competent, or just a cheat? Interactive effects of honesty-humility and moral cues on cheating behavior. *Personality and Social Psychology Bulletin*, 44(2), 158–172. DOI: 10.1177/0146167217733071.
- Krosnick, J.A., and Petty, R.E. (1995). Attitude strength: An overview. In R.E. Petty and J.A. Krosnick (eds), *Attitude strength: Antecedents and consequences* (pp. 1–24). Erlbaum.
- Lammers, J., Stapel, D.A., and Galinsky, A.D. (2010). Power increases hypocrisy: Moralizing in reasoning, immorality in behavior. *Psychological Science*, 21(5), 737–744. DOI: 10.1177/095679761036881.
- Lamprinakos, G., Magrizos, S., Kostopoulos, I., Drossos, D., and Santos, D. (2022). Overt and covert customer data collection in online personalized advertising: the role of user emotions. *Journal of Business Research*, 141, 308–320. DOI: 10.1016/j.jbusres.2021.12.025.
- Lamprinakos, G., Stavraki, M., Briñol, P., and Petty, R.E. (2024a). Moral emotions influence information processing and persuasion: A differential appraisal perspective. (Manuscript under Review.)
- Lamprinakos, G., Stavraki, M., Santos, D., Briñol, P., and Petty, R.E. (2024b). Power can increase but also decrease cheating depending on the thoughts validated. *Journal of Experimental Social Psychology*, 111, ArtID:104578. DOI:10.1016/j.jesp.2023.104578.
- Lee, S.W.S., Chen, K., Ma, C., and Hoang, J. (2024). Wipe it off: A meta-analytic review of the psychological consequences and antecedents of physical cleansing. *Psychological Bulletin*, 150(4), 355–398. DOI: 10.1037/bul0000421.
- Lee, S.W.S., and Schwarz, N. (2011). Wiping the slate clean: Psychological consequences of physical cleansing. *Current Directions in Psychological Science*, 20, 307–311. DOI: 10.1177/0963721411422694.
- Lee, S.W.S., and Schwarz, N. (2020). Grounded procedures: A proximate mechanism for the psychology of cleansing and other physical actions. *Behavioral and Brain Sciences*, 1–78. DOI: 10.1017/S0140525X20000308.
- Lerner, J.S., and Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition and Emotion*, 14(4), 473–493. DOI: 10.1080/026999300402763.
- Luttrell, A. (forthcoming). Appealing to morality and values: A personalized matching account. In R.E. Petty, A. Luttrell, and J.D. Teeny (eds). *The handbook of personalized persuasion: Theory and application*. Taylor & Francis.
- Luttrell, A., and Petty, R.E. (2021). Evaluations of self-focused versus other-focused arguments for social distancing: An extension of moral matching effects. *Social Psychological and Personality Science* 12(6), 946–954. DOI: 10.1177/1948550620947853.
- Luttrell, A., Petty, R.E., Briñol, P., and Wagner, B.C. (2016). Making it moral: Merely labelling an attitude as moral increases its strength. *Journal of Experimental Social Psychology*, 65, 82–93. <https://doi.org/10.1016/j.jesp.2016.04.003>.
- Luttrell, A., Phillip-Muller, A., and Petty, R.E. (2019). Challenging moral attitudes with moral messages. *Psychological Science*, 30(8) 1136–1150. <https://doi.org/10.1177/0956797619854706>.
- Luttrell, A., and Sawicki, V. (2020). Attitude strength: Distinguishing predictors versus defining features. *Social and Personality Psychology Compass*, 14, e12555. DOI: 10.1111/spc3.12555.

- Luttrell, A., and Togans, L.J. (2021). The stability of moralized attitudes over time. *Personality and Social Psychology Bulletin*, 47(4), 551–564. DOI: 10.1177/0146167220935737.
- Minson, J.A., and Chen, F.S. (2022). Receptiveness to opposing views: Conceptualization and integrative review. *Personality and Social Psychology Review*, 26(2), 93–111. DOI: 10.1177/10888683211061037.
- Moreno, L., Requero, B., Santos, D., Paredes, B., Briñol, P., and Petty, R.E. (2021). Attitudes and attitude certainty guiding pro-social behavior as a function of perceived elaboration. *European Journal of Social Psychology*, 51, 990–1006. DOI: 10.1002/ejsp.2798.
- Paredes, B., Briñol, P., Petty, R.E., and Gómez, A. (2021). Increasing the predictive validity of identity fusion in leading to sacrifice by considering the extremity of the situation. *European Journal of Social Psychology*, 51, 239–250. DOI:10.1002/ejsp.2727.
- Paredes, B., Santos, D., Briñol, P., Gómez, A., and Petty, R.E. (2020). The role of meta-cognitive certainty on the relationship between identity fusion and endorsement of extreme pro-group behavior. *Self and Identity*, 19(7), 804–824.
- Paredes, B., Stavradi, M., Briñol, P., and Petty, R.E. (2013). Smiling after thinking increases reliance on thoughts. *Social Psychology*, 44, 349–353. DOI: 10.1027/1864-9335/a000131.
- Paulhus, D.L., and Williams, K.M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, 36(6), 556–563. DOI: 10.1016/S0092-6566(02)00505-6.
- Petty, R.E., Briñol, P., Fabrigar, L.R., and Wegener, D.T. (2019). Attitude structure and change. In R.F. Baumeister and E.J. Finkel (eds), *Advanced social psychology* (2nd edn, pp. 117–156). Oxford University Press.
- Petty, R.E., Briñol, P., and Tormala, Z.L. (2002). Thought confidence as a determinant of persuasion: The self-validation hypothesis. *Journal of Personality and Social Psychology*, 82, 722–741. DOI: 10.1037/0022-3514.82.5.722.
- Petty, R.E., Briñol, P., Tormala, Z.L., and Wegener, D.T. (2007). The role of meta-cognition in social judgment. In E.T. Higgins and A.W. Kruglanski (eds), *Social psychology: A handbook of basic principles* (2nd edn, pp. 254–284). Guilford Press.
- Petty, R.E., and Krosnick, J.A. (1995). *Attitude strength: Antecedents and consequences*. Erlbaum Associates.
- Petty, R.E., Luttrell, A., and Teeny, J.D. (eds) (forthcoming). *The handbook of personalized persuasion: Theory and application*. Taylor & Francis.
- Petty, R.E., Siev, J.J., and Briñol, P. (2023). Attitude strength: What's new. *Spanish Journal of Psychology*, 26, e4, 1–13. DOI: 10.1017/SJP.2023.7.
- Pfafftheicher, S., Schindler, S., and Nockur, L. (2019). On the impact of honesty-humility and a cue of being watched on cheating behavior. *Journal of Economic Psychology*, 71, 159–174. DOI: 10.1016/j.joep.2018.06.004.
- Requero, B., Santos, D., Paredes, B., Briñol, P., and Petty, R.E. (2020). Attitudes toward hiring people with disabilities: A meta-cognitive approach to persuasion. *Journal of Applied Social Psychology*, 50(5), 276–288. DOI: 10.1111/jasp.12658.
- Santos, D., Briñol, P., Petty, R.E., Gandarillas, B., and Mateos, R. (2019). Trait aggressiveness predicting aggressive behavior: The moderating role of meta-cognitive certainty. *Aggressive Behavior*, 45(3), 255–264. DOI: 10.1002/ab.21815.
- Santos, D., Ghodsinia, A., Requero, B., Gonçalves, D., Briñol, P., and Petty, R.E. (2024). Certainty improves the predictive validity of Honesty-Humility and Dark Triad traits on cheating behavior. (Manuscript under review.)
- Santos, D., Paredes, B., Briñol, P., and Petty, R.E. (2022). Trait aggressiveness and aggressive behavior: The simultaneous influence of contextual variables. *Psychology of Violence*, 12(6), 438–449. DOI: 10.1037/vio0000446.
- Santos, D., and Rivera, R.G. (2015). The accessibility of justice-related concepts can validate intentions to punish. *Social Influence*, 10(3), 180–92. DOI: 10.1080/15534510.2015.1031170.
- Shoots-Reinhard, B., Petty, R.E., DeMarree, K.G., and Rucker, D.D. (2015). Personality certainty and politics: Increasing the predictive utility of individual difference inventories. *Political Psychology*, 36, 415–430. DOI: 10.1111/pops.12104.

- Skitka, L.J., Bauman, C.W., and Sargis, E.G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, 88, 895–917. DOI: 10.1037/0022-3514.88.6.895.
- Skitka, L.J., Hanson, B.E., Morgan G.S., and Wisneski, D.C. (2021). The psychology of moral conviction. *Annual Review of Psychology*, 72, 347–366. DOI: 10.1146/annurev-psych-063020-030612.
- Snyder, M., and DeBono, K.G. (1985). Appeals to image and claims about quality: Understanding the psychology of advertising. *Journal of Personality and Social Psychology*, 49, 586–597. DOI: 10.1037/0022-3514.49.3.586.
- Suri, S., Goldstein, D.G., and Mason, W.A. (2011). Honesty in an online labor market. In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence, Human Computation 2011*, 61–66.
- Swann Jr, W.B., Gómez, Á., Buhrmester, M.D., López-Rodríguez, L., Jiménez, J., and Vázquez, A. (2014). Contemplating the ultimate sacrifice: Identity fusion channels pro-group affect, cognition, and moral decision making. *Journal of Personality and Social Psychology*, 106, 713–727. DOI: 10.1037/a0035809.
- Teeny, J.D., Siev, J.J., Briñol, P., and Petty, R.E. (2021). A review and conceptual framework for understanding personalized matching effects in persuasion. *Journal of Consumer Psychology*, 31(2) 382–414. DOI: 10.1002/jcpsy.1198.
- Toader, I., Moreno, L., Briñol, P., and Petty, R.E. (2024). Power corrupts and absolute power (being sure of felt power) corrupts absolutely. (Manuscript under review.)
- Tormala, Z.L., Falces, C., Briñol, P., and Petty, R.E. (2007). Ease of retrieval effects in social judgment: The role of unrequested cognitions. *Journal of Personality and Social Psychology*, 93, 143–157.
- Wichman, A.L., Briñol, P., Petty, R.E., Rucker, D.D., Tormala, Z.L., and Weary, G. (2010). Doubting one's doubt: A formula for confidence. *Journal of Experimental Social Psychology*, 46, 350–355. DOI: 10.1016/j.jesp.2009.10.012.
- Xu, M., and Petty, R.E. (2022). Two-sided messages promote openness for morally based attitudes. *Personality and Social Psychology Bulletin*, 48(8), 1151–1166. DOI: 10.1177/0146167220988371.
- Xu, M., and Petty, R.E. (2024). Two-sided messages promote openness for a variety of deeply entrenched attitudes. *Personality and Social Psychology Bulletin*, 50(2), 215–231. DOI:10.1177/01461672221128113.
- Zhong, B., and Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science*, 313, 1451–1452. DOI: 10.1126/science.1130726.