# An Appraisal Model of Criminal Decision Making

## How Person Factors Affect Decisions through Cognitive Appraisals

Inaugural-Dissertation zur Erlangung der Doktorwürde

der

Philosophischen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität

zu Bonn

vorgelegt von

**Anja Carina Murmann**

aus

Geldern

Bonn, 2021

Gedruckt mit der Genehmigung der Philosophischen Fakultät
Der Rheinischen Friedrich-Wilhelms-Universität Bonn

**Zusammensetzung der Prüfungskommission:**

Prof. Dr. Ulrich Ettinger, Institut für Psychologie, Universität Bonn
*(Vorsitzender)*

Prof. Dr. Rainer Banse, Institut für Psychologie, Universität Bonn
*(Betreuer und Gutachter)*

Prof. Dr. Andreas Mokros, Fakultät für Psychologie, FernUniversität Hagen
*(Gutachter)*

PD Dr. Ina Grau, Institut für Psychologie, Universität Bonn
*(Weiteres prüfungsberechtigtes Mitglied)*

Tag der mündlichen Prüfung: 15.01.2021

**Danksagung / Acknowledgement**

Allen voran bedanke ich mich bei dir, Rainer. Als ich für meinen Master nach Bonn kam, um mehr über das Fach Rechtspsychologie zu lernen, hast du mir die Möglichkeit gegeben, als studentische Hilfskraft in deiner Abteilung Fuß zu fassen und mich später ermutigt, eine Promotion zu beginnen. Du hast mich stets gefördert und mir dein Vertrauen entgegengebracht. Vielen Dank für alles, was ich von dir gelernt habe – im Bereich Rechtspsychologie, aber auch darüber hinaus. Ich danke dir außerdem für die Projektidee und die Beratung und Unterstützung bei der Modell-Entwicklung sowie bei allen weiteren Phasen meiner Promotion.

An dieser Stelle möchte ich mich auch bei allen Studierenden und studentischen Hilfskräften bedanken, die zu dem Erfolg dieses Projektes beigetragen haben. Ich danke

- Elisa Krause und den Studierenden des Projektarbeits-Kurses 2017/18 für die Beteiligung an der Datenerhebung der Studien 1 und 2
- Alexandra Große und Nathalie Losigkeit für die Beteiligung an der Generierung von Vignettenmaterial und an der Planung und Datenerhebung der Studie 3
- Felix Ter-Nedden für die Beteiligung an der Entwicklung des Leitfadens und an der Durchführung der Interviews in Studie 4
- Philipp Musfeld, der mich nicht nur bei der Auswertung der Studien 1-3 unterstützt hat, sondern maßgeblich an der Konzeption und Analyse der Studie 5 beteiligt war. Vielen Dank, Philipp, für deine hilfreichen Impulse bei der Auswertung der Studie und bei der Weiterentwicklung des Modells!

Ich bedanke mich auch bei meinen Kolleginnen Charis, Michaela, Laura, Verena, Lisa, Christine und Ina. Es hat großen Spaß gemacht, mit euch zusammenzuarbeiten (von den Tagungen ganz zu schweigen)! Danke auch an dich, Carolin, für die Unterstützung in allen organisatorischen Belangen.

Zuletzt danke ich meinen Freundinnen und Freunden und meiner Familie; insbesondere meinen Eltern Elisabeth und Albrecht und meinem Bruder Thilo, dafür, dass ihr mich in jeder Lebenslage unterstützt. Ich danke dir, Niklas, dass du auch in stressigen Zeiten immer für mich da bist.

## Abstract

In explaining and predicting criminal behavior, research has mainly concentrated on stable person factors such as psychopathy or self-control. However, the causal processes underlying these correlates of crime are largely unknown. In the present research, the Appraisal Model of Criminal Decision Making is introduced. The model postulates the dimensions *Incentive*, *Feasibility*, *Legality*, *Morality*, *Likelihood*, and *Level of Punishment* as crucial cognitive mechanisms in delinquent decisions. Five studies investigated whether the dimensions affect decision making in general, and whether they interact with established criminogenic risk factors. Participants were presented with vignettes that were manipulated regarding the appraisal dimensions and asked whether they would execute the respective behaviors. The vignettes described low-level criminogenic opportunities in Studies 1 and 2 ($N$ = 299), and dissexual behavior opportunities in Study 3 ($N$ = 685). All three studies showed that the appraisal dimensions influenced decisions. However, the *Incentive* dimension did not affect decisions in Studies 1 and 2, which might be explained by a hot-cold-empathy gap. In both behavior domains, the manipulation of *Morality* had the strongest influence on decision making. Additionally, there were indications that personality factors (psychopathy, morality, and self-control) interacted with appraisal dimensions. In Study 4, former offenders ($N$ = 22) were interviewed about appraisal processes in the moment of decision making. All of the appraisal dimensions were named; however, which dimensions were named depended highly on the context. Some participants found illegal actions particularly appealing, which suggests that the *Legality* dimension might interact with personality factors. In Study 5, a regression-analytical approach was chosen to overcome the methodological problems from Studies 1-3. In a pre-study, an independent sample was asked to rate all vignettes regarding the appraisal dimensions. The normative values generated in the pre-study were analyzed in combination with the decision values of the main study participants ($N$ = 198). All dimensions except *Feasibility* affected decision making. Furthermore, participants high in psychopathy weighted the *Morality* and *Legality* dimensions less strongly, and participants high in sensation-seeking weighted the *Likelihood of Punishment* and *Legality* dimensions less strongly. These results indicate that appraisal processes provide a causal link between established criminogenic personality factors and criminal behavior and thus contribute to a better understanding of crime.

**Table of Contents**

## Introduction

When asked why he robbed banks, the infamous bank robber Willie Sutton (1901-1980) is said to have answered: “Because that’s where the money is” (Cocheo, 1997; Cornish & Clarke, 2011). As plausible as this answer seems, it is questionable whether human decision making is really that simple. The question of why people show criminal behavior has concerned researchers all over the world, along with the questions how to predict and prevent crime. In predicting criminal behavior, criminology and legal psychology have mainly focused on relatively stable person factors or traits such as gender (Broidy & Agnew, 1997; Newburn & Stanko, 1994), age (Farrington, 1986; Moffitt, 1993), antisocial personality (Andrews & Bonta, 2010), psychopathy (Hare & Neumann, 2009; Hart & Hare, 1997), or self-control deficits (Gottfredson & Hirschi, 1990). Although this approach has worked rather well in predicting future delinquency, these correlates have little explanatory value for criminal behavior.

Risk assessment instruments that are used to predict the probability of future criminal behavior are mainly based on past behavior (e.g., actuarial risk measures) or on personality traits assessed through self-report measures tapping into the respondent’s representation of past behavior (e.g., aggressiveness), or attitudes and values (e.g., pro-criminal attitudes) that are very much anchored in past behavior. In large part, the antisocial personality is inferred from past antisocial behavior and used to predict future antisocial behavior. But common risk factors such as antisocial personality, psychopathy, and self-control are of limited value when it comes to understanding and explaining criminal behavior, as the mediating psychological processes causing criminal behavior in a specific situation are not considered.

However, in order to develop strategies for the prevention and treatment of criminal behavior, it is essential to understand these mediating processes (Miller & Lynam, 2001; van Gelder & Vries, 2012). We[1] propose that one decisive underlying the correlation between person factors

[1] The model was developed in cooperation with the supervisor of this thesis and all five studies were conducted in collaboration with students. Therefore, I refer to "we" when describing the model and reporting the studies.

and crime is appraisal of the criminogenic situation. Therefore, a new approach is presented in which these cognitive appraisal processes within a criminogenic situation are analyzed alongside established criminogenic personality traits. More specifically, we investigate what it means to, for example, have a psychopathic personality pattern with respect to the appraisal of certain features of a criminogenic situation and subsequent criminal or dissocial behavior. Which appraisal processes generally take place within a criminogenic situation in most people? How do the appraisals of a situation determine whether an opportunity for criminal action is chosen or rejected? And to what extent are appraisals influenced by stable and dynamic person factors?

In this thesis the Appraisal Model of Criminal Decision Making, that is akin to appraisal theories of emotion (Scherer, 1997, 2009), theories of prosocial behavior (Latané & Darley, 1970) and established criminological theories, is presented together with five studies to test the validity of the model. Before presenting the model, I briefly summarize the criminological theories on which our model is based.

**Previous Theoretical Approaches**

The first approaches that have addressed situational conditions of criminal behavior are opportunity approaches, such as the Routine Activity Theory (RAT; Cohen & Felson, 1979), and rational choice approaches (Cornish & Clarke, 1986, 2011). The RAT concentrates on circumstances in which people carry out criminal acts and is therefore able to explain changes in crime rates. It postulates that criminal acts require convergence in space and time of *likely offenders*, *suitable targets,* and the absence of *capable guardians* against crime (Cohen & Felson, 1979). Thus, both the characteristics of the perpetrator and the characteristics of the situation in terms of the attractiveness of the target and the existence of control mechanisms play a role in this theory. However, since RAT is a macro theory of crime, it cannot explain what constitutes a motivated offender and which cognitive mechanisms determine whether an opportunity for criminal behavior is seized or not.

Rational choice approaches, on the other hand, do consider cognitive mechanisms, because they conceptualize criminal behavior as the result of a conscious and rational decision-making process (Cornish & Clarke, 1986). Accordingly, criminal acts occur when the subjectively perceived utility and probability of success have a higher balance against the negative

consequences as compared to non-criminal alternatives. However, deterrence research points out that in this equation the risk of negative consequences is more relevant than the severity of the sanction that may be imposed (Lösel & Schmucker, 2008). A large number of studies have found a consistent deterrent effect of perceived certainty of sanctions (e.g., Cole, 1989; Klepper & Nagin, 1989; Kraut, 1976; Maxwell & Gray, 2007; Tittle & Rowe, 1974), whereas evidence for the deterrent influence of perceived severity of sanctions is often lacking (e.g., Silberman, 1976; Waldo & Chiricos, 1972).

Meanwhile, rational choice theory has become a more comprehensive perspective that acknowledges the complexity of criminal behavior (Pratt, 2008). For example, it has been further specified in terms of which elements are particularly decisive for which individuals. Based on data from the longitudinal Dunedin birth cohort study of individuals to the age of 26 ($N$ = 1,002), Wright, Caspi, Moffitt, and Paternoster (2004) found that the view that criminal behavior is costly and risky most deterred people with low self-control and high levels of self-perceived criminality. However, other studies found that deterrence only worked for people with high self-control (Nagin & Pogarsky, 2001; Piquero & Tibbetts, 1996; Pogarsky, 2002). Beyond that, there is evidence that the way people perceive costs and benefits of engaging in criminal behavior also depends on other individual and contextual factors, such as prior experiences with crime and punishment, attachment to prosocial institutions, and environmental constraints for criminal opportunities (see Pratt, 2008 for an overview).

Nevertheless, there is criticism against the theory that too little consideration is given to interindividual differences (Lösel & Schmucker, 2008). The rational choice theory sometimes appears limited and unrealistic in portraying criminal decision processes and takes too little account of affective aspects (De Haan & Vos, 2003; van Gelder, 2013). A meta-analysis on the empirical status of deterrence theory has shown that the effect sizes of detection probability and especially punishment severity were only small to medium (Pratt et al., 2006). Piliavin et al. (1986) conclude that the rational choice theory oversimplifies the cognitive processes behind criminality and call for a more complex model.

A more recent theory that sees criminal behavior as the result of the interplay between individual and environmental factors is the Situational Action Theory (SAT; Wikström, 2004, 2006; Wikström & Treiber, 2009). It integrates personality traits and contextual factors such as culture and community. In a nutshell, this means, “People do what they do because of who

they are and the features of the environments in which they take part" (Wikström, 2014, p. 75). According to the SAT, a criminal act is an outcome of a perception-choice process. The perception-choice process is initiated and guided by relevant aspects of the person-environment interaction. As with the RAT, a motivation (e.g., temptation, provocation) of the perpetrator is necessary, but not sufficient for committing criminal acts. A crime will only be committed if the person perceives the criminal act as a possible alternative. Whether this is the case or not depends on the person's moral values (*moral filter*). If the criminal action is perceived as a viable alternative, moral habits may emerge that trigger an automatic response (i.e., criminal behavior) to a habitual situation. If there is no habit, deliberation will take place. Now the decision will depend on the effectiveness of internal (self-control) and external (deterrents such as CCTV cameras or presence of police officers) controls. Controls only come into play when the moral filter has not succeeded in excluding crime from the variety of perceived action alternatives. For most people, in most circumstances, the ability to exert self-control should be irrelevant for their course of action, because they either act habitually or, due to their moral filter, do not see crime as an alternative (Wikström & Treiber, 2007).

Wikström and Treiber (2009) also describe that theories often fail to distinguish between correlates or markers and actual causes of crime, because an understanding of the causal processes is lacking. This fundamental consideration is the starting point for our theory. The SAT is understood as a general theory of moral action that aims to explain why people break moral rules and in which crime is regarded as a subclass of acts of moral rule breaking (Wikström & Treiber, 2009). According to the authors, there is no fundamental difference between explaining why people break moral rules in general (e.g., drinking alcohol before noon) and why they break moral rules defined by law. The causal processes are considered the same.

In our view, appraisals of legality and morality constitute unique dimensions that both affect decision making. Actions can be avoided because they are immoral without being illegal and vice versa. Additionally, in the SAT, no specific assumptions are made on which processes ultimately lead to a person choosing a certain behavioral option. The SAT is less focused on the situation itself, as it includes the effects of habits and traits such as self-control without translating them into cognitive mechanisms. The model presented in the following section

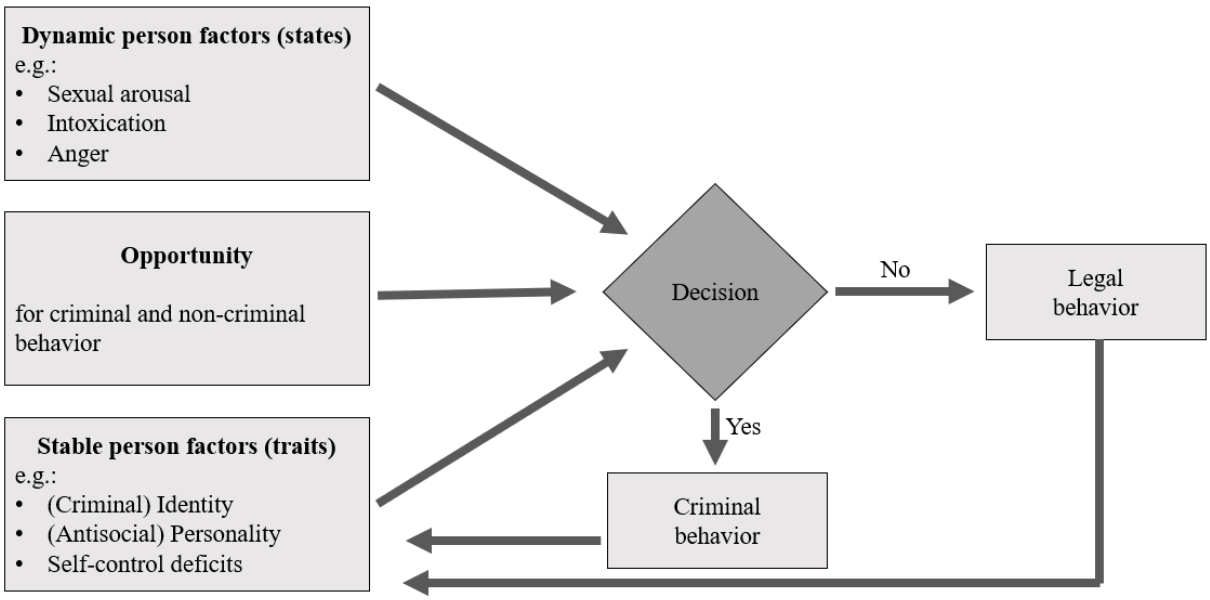
addresses these questions by explaining how the appraisal of a situation determines whether an illegal behavior option is chosen or not.

## The Appraisal Model of Criminal Decision Making

According to the framework model, dynamic person factors (states), stable person factors (traits), and opportunities (situations) play together and influence whether a person ultimately decides in favor of or against a criminal action (Figure 1). It integrates established theories and risk factors of criminal behavior (i.e., stable and dynamic person factors) and the behavioral decision approach (Figure 1).

**Figure 1.**

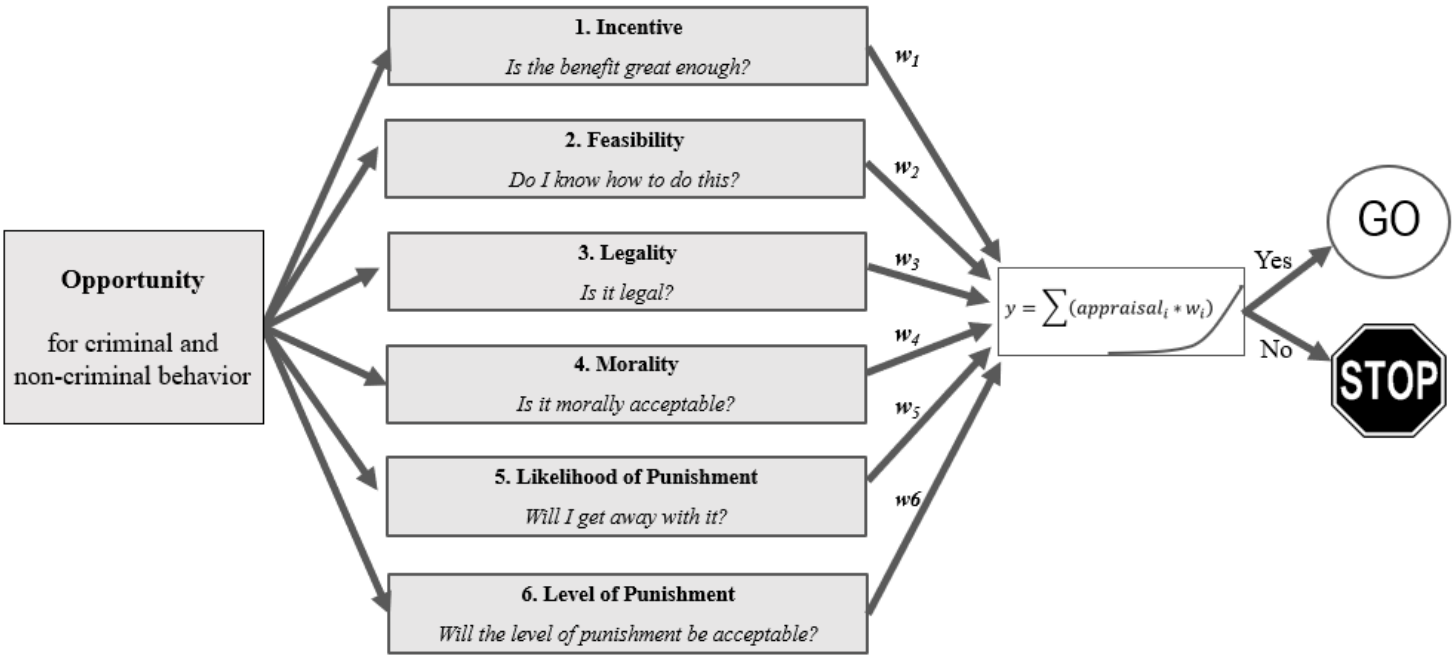
*Framework Model of Criminal Decision Making*




We assume that the causal link between these person factors and criminal decision making consists of cognitive appraisals (Figure 2). Depending on appraisals, a person will either decide in favor of or against a criminal action. The behavioral outcomes, in turn, can affect traits. For example, showing criminal behavior can strengthen one's criminal identity (Asencio & Burke, 2011) or weaken one's beliefs about self-control capacities (Job et al., 2010), which can result in different appraisals of new situations and therefore in a higher frequency of illegal decisions.

**Figure 2**

*Appraisal Model of Criminal Decision Making*




We introduce the concept of appraisals that originates from emotion research (Scherer, 1997, 2009) to the domain of criminal behavior research. According to appraisal theories of emotion, whether a certain event evokes an emotion in a person -- and if so, which emotion and with what intensity -- depends on how the event is appraised by this person. In line with that, we propose that criminogenic situations are appraised with regard to distinct appraisal dimensions, and that a behavioral decision is a result of these appraisal processes. Our appraisal model (Figure 2) is also akin to behavioral decision models such as the emergency intervention process model by Latané and Darley (1970).

As in these models we initially assumed a fixed sequential order of the different appraisal dimensions and were interested in analyzing this sequential order. However, after conducting the first three studies we realized that for the mainly differential question of who displays criminal behavior in a certain situation and who does not, the serial or parallel architecture of the appraisal model is of minor importance. The critical features of the model are the number and kinds of appraisal dimensions and how they are processed in order to make a behavioral decision.

We postulate that each criminogenic situation is appraised in terms of six dimensions: 1) *Incentive*: Is the benefit great enough? 2) *Feasibility*: Do I know how to do this? 3) *Legality*: Is it legal? 4) *Morality*: Is it morally acceptable? 5) *Likelihood of Punishment*: Will I get away with it? 6) *Level of Punishment*: Will the level of (formal and informal) punishment be acceptable? Of course, one does not actually ask such a series of questions each time when appraising a situation; appraisal is not an internal dialogue (Ellsworth & Scherer, 2003). The appraisals can be made at a conscious level, however, they can also be unconscious (Scherer, 2005).

Each of the six dimensions is assigned a weight that reflects the relevance of the appraisal dimension for the behavioral decision, depending on stable dispositions and on the current state of the person. After the weights are assigned, the weighted appraisals are added together. The larger the sum, the more likely the behavior will be executed. In real life, the decision is binary (yes or no), and the behavior will be executed if a certain threshold is exceeded.

We propose that stable and dynamic personal factors (traits and states) can influence this decision-making process in two ways: (1) by influencing the weight (*w*) that is assigned to the dimension, and (2) by influencing the appraisal itself due to different individual setpoints.

(1) The weight of an appraisal dimension is equivalent to the personal relevance of this dimension, i.e., the extent to which this dimension is considered in the decision. If a dimension is appraised negatively but the corresponding dimension is of little relevance to this person in this specific situation (i.e., it has a low weight), it is likely that the behavior will still be executed. Dimensions can have such a low weight that the appraisal of the situation with regard to this dimension is practically skipped. For example, a person high in psychopathy might appraise a certain behavior as immoral but still execute the behavior because the weight of the *Morality* dimension is very low, and the other appraisals that speak in favor of the action are weighted more strongly. Or a person low in self-control might weight incentives so strongly that it becomes very difficult to resist the temptation, even though other appraisals are negative. If a person weights the legality check strongly, the mere fact that the behavior violates the law can be decisive for the behavior not being shown – even if there is no other reason against it.

(2) We assume that not only the weighting, but also the appraisal itself varies depending on the traits and states of a person. Someone with a high psychopathy score may consider a certain behavior as not immoral at all, whereas someone scoring low in psychopathy would appraise it as highly immoral. This distorted appraisal is caused by a different individual moral setpoint, i.e., everything below this point is considered morally acceptable. This setpoint metaphor stems from emotion research that showed that people with depression have an inappropriate setting of the neutral point of the hedonic detection system, with the result that previously neutral stimuli appear negative (Baddeley et al., 2012). We postulate that there are similar setpoints for the appraisal dimensions. The setpoint can be shifted not only by stable personality factors, but also by dynamic factors, such as arousal or intoxication. A drunk person might underestimate the likelihood and/or the severity of a potential punishment, and a sexually aroused person might appraise even weak erotic stimuli as very attractive (Ariely & Loewenstein, 2006).

The model proposed here is the simplest possible model of this kind. It is possible to develop more complex models. For example, a more complex model could be a single or multiple cutoff model, according to which individual dimensions must reach a certain cutoff in order for the behavior to be executed. If this cutoff is not reached, this would function as a veto (e.g., if something is appraised as illegal, it is not considered an option). The appraisals of the other dimensions would not be able to compensate for this. In the first studies of this thesis the focus was on determining whether all appraisal dimensions play a role in decision making. In Study 5, we used a comprehensive analysis to test the simple additive appraisal model described above.

The main potential of our approach is to show that criminogenic person factors such as psychopathy or self-control are reflected in certain appraisal patterns. This makes it possible to investigate the causal processes underlying the correlations between these factors and crime. If people change regarding their personality or attitudes, e.g., in the process of desistance from crime (Laub & Sampson, 2001), these changes might be mapped in form of appraisals. This could open up new perspectives for research and practice (see General Conclusion and Outlook).

**Previous Studies Using the Scenario Method**

In a first step, we examined whether the postulated dimensions have an impact on the intention to show criminal behavior. Therefore, we presented participants with small vignettes depicting opportunities for criminal behavior (i.e., hypothetical scenarios) and asked them if they would show this behavior or not. Vignettes offer the possibility to expose participants to a large variety of situations and to manipulate single aspects of these situations, which would be impossible in real situations. Although behavioral intentions are not synonymous with actual behavior, they are highly correlated and provide good estimates of actual behavior (Fishbein & Ajzen, 1975; Green, 1989; Kim & Hunter, 1993; Murray & Erickson, 1987; Pogarsky, 2004).

There have been several attempts to study the impact of some of the dimensions that are also addressed in our theoretical model of criminal decisions. In most cases, the studies were concerned with the benefits and formal costs of an action, because they referred to rational choice or deterrence theories (Bachman et al., 1992; Klepper & Nagin, 1989; Paternoster & Simpson, 1996), or with social sanctions, because they referred to social control theories (Nagin & Paternoster, 1994). However, often only one specific type of crime was addressed; for example, Thurman et al. (1993) showed that the decision to drink and drive was influenced by factors like weather conditions, number of miles to drive, legal consequences, community response, etc. Paternoster and Simpson (1996) found that the decision to commit corporate crimes was affected by sanction threats, moral evaluations, and organizational factors. In the same study, fear of sanctions had no effect when respondents were inhibited by their moral evaluation, which could be taken as an indication of the effectiveness of a moral filter as conceptualized by Wikström (2010). The same effect was also found by Bachman et al. (1992) in their study that contained vignettes describing a sexual assault.

A vignette study by Nagin and Paternoster (1993), which linked time-stable individual differences in propensity to offend with situational factors for the first time, showed that both "criminal propensity" (operationalized as self-control, measured by the 24-item scale by Grasmick et al., 1993) and proximate situational influences (e.g., the attractiveness of the crime target, the ease of committing the crime with minimum risk, perceptions of costs and benefits of the crime) affected criminal decision making. Self-control was directly and indirectly related to intentions to offend, i.e., through its influence on choice-relevant

variables: Persons low in self-control perceived the rewards of crime as more valuable and the costs of crime as less aversive, were less likely to feel ashamed, and showed stronger intentions to commit crimes than people with more self-control.

Van Gelder and de Vries (2012, 2014) extended this research that links stable individual difference factors (e.g., self-control) to proximal states in the moment of decision making and introduced the "Hot/Cool Perspective of Criminal Decision Making" (van Gelder, 2013). They proposed two different mental processing modes, a "hot" affective mode and a "cool" cognitive mode and showed that individual difference factors, such as self-control, were directly and indirectly related to criminal decision making. This indirect relation was mediated by affect, i.e., feelings of fear and worry, and rational choice variables, i.e., perceived risk of sanction, which represents the two processing modes. The authors argued that decision making is not only influenced by the anticipation of negative affects but also by the affective states experienced at the time of the decision (see also Carmichael & Piquero, 2004). These ideas are in accordance with the assumptions of our appraisal model because the anticipation of negative affects is reflected in the dimensions *Morality* (considering an action as immoral can result in anticipating feelings of shame or guilt) and *Likelihood / Level of Punishment* (punishment can result in negative emotions). Emotions can either influence appraisals (e.g., anger), or can arise from appraisals (e.g., anticipated negative or positive emotions). Therefore, our approach offers the advantage that these factors are addressed and analyzed on only one psychological level (i.e., appraisals).

From a methodological point of view, it can be criticized that in previous studies participants were only presented with a very small number of scenarios, which limits external validity. Additionally, it cannot be ruled out that findings can be explained by the participants' striving for consistency. Participants were asked to estimate the probability that they would commit the act specified in the vignette, but also to answer questions regarding the choice-relevant variables, such as the probability of costs and benefits or anticipated emotions (Bachman et al., 1992; Nagin & Paternoster, 1993; Paternoster & Simpson, 1996; van Gelder & Vries, 2012, 2014), thus, these estimations may have affected each other.

In the present research, we pursue the link between theories of time-stable criminal propensity, and theories of criminal opportunity in greater detail, with a more comprehensive

theoretical approach that combines existing theories. We use a considerably larger number of vignettes, and vignettes that are more contemporary than in previous studies.

**Overview of the Present Studies**

In the first part of this dissertation, three studies are presented that test the key assumption of the model, i.e. the dimensions' influence on decision making. Studies 1-3 use an experimental design and focus on whether the dimensions have any influence on decision making at all. In addition, first indications of interactions between the dimensions and trait variables are examined. In studies 4-5, the primary focus is on investigating these interactions with established risk factors for delinquency. Study 4 uses a qualitative approach in the form of interviews with former offenders; Study 5 uses a regression-analytical approach and takes up the scenario method that was also used in Studies 1-3. In the following first part of the thesis, Studies 1-3 are described, and, after an interim discussion, Studies 4 and 5 follow in the second part.

The first two studies deal with low-level crimes of university students while the third study is on dissexual behavior and sexual offenses in an all-male sample. In all three studies we used vignettes that offered opportunities to break the law. Participants made hypothetical decisions as to whether they would show the critical behaviors or not. Within the vignettes, we manipulated the appraisal dimensions *Incentive*, *Legality*, *Morality*, *Likelihood of Punishment*, and *Level of Punishment*, and tested whether these manipulations had an effect on behavior decisions. Feasibility was not manipulated because we considered its effect on decision making as trivial.

To ensure that the participants could identify with the scenarios and to enhance ecological validity, we designed them as realistic as possible. Because in a real-life criminogenic situation the decision must be either yes or no, we opted for a binary decision category. A recent study by Waubert de Puiseau et al. (2019) using both probability ratings and binary decision categories showed that both measures were highly correlated ($r$ = .82). In contrast to other studies that used vignettes dealing with criminal actions of another person (e.g. (Bachman et al., 1992; Klepper & Nagin, 1989; Paternoster & Simpson, 1996), we chose to address the participants directly in the second person singular ("you").

Our initial idea was to investigate whether the appraisal dimensions are run through in a certain sequence. To draw conclusions about this sequence we recorded participants' response times (RTs). In hindsight, we now consider the analysis of the RTs to be of only limited informative value because the lengths of the RTs cannot provide direct information about the number of appraisals. Even though most RT hypotheses have been formally confirmed, the preregistered RT hypotheses and the corresponding results are not reported here. Explanation and discussion of why we do not consider RTs informative, hypotheses and results of the analyses regarding RTs can be found in Appendix A.

## Studies 1 and 2

In Studies 1 and 2 we investigated if the appraisal dimensions had an influence on daily criminal decision making in student samples. Our first hypothesis was that a manipulation of each of the five dimensions of the appraisal model (*Incentive*, *Legality*, *Morality*, *Likelihood of Punishment*, *Level of Punishment*) within the vignettes would have an impact on the frequency of positive decisions.

In addition to criminal decision making, we assessed the following person variables using self-report questionnaires: Big Five personality traits, psychopathy, demographics, and criminal lifestyle information (convictions, drug consumption). Big Five personality traits were assessed for exploratory purposes, and we chose to assess psychopathy because it has been identified as one of the strongest individual-level predictors of general offending (DeLisi, 2009) besides age and gender. Therefore, we expected participants with a high psychopathy score to show a higher frequency of positive responses for criminal behavior than participants low in psychopathy (Hypothesis 2), and to react in a less sensitive way to the manipulation of appraisal dimensions (Hypothesis 3).

In the first study, the vignettes on everyday delinquency were presented in written form, whereas in the second study, the same vignettes were presented acoustically. We chose this acoustic presentation mode in Study 2 in order to ensure that the participants captured all the vignette information instead of skipping parts of vignette description. The hypotheses, procedure, and materials were identical in both studies. Because the results were also very

similar, both studies were combined and the results from an overall analysis will be reported. The studies were preregistered in the Open Science Framework (Appendix B).[2]

## Method Study 1

### *Participants*

A total of 100 university students and recent alumni (78 women, 22 men) took part in this online study. We opted for this sample for reasons of convenience and because student samples are likely to contain moderate numbers of offenders (Nagin & Paternoster, 1993). Psychology students received partial course credit in exchange for their participation. Participants were between the ages of 18 and 41 ($M$ = 23.89, $SD$ = 3.13). A total of 84% of participants reported having consumed alcohol at least once in the last 30 days (tobacco: 26%, cannabis: 14%, hard drugs: 6%). Two participants claimed to have a criminal record.

### *Materials and Design*

**Vignettes.** We used vignettes depicting typical criminogenic situations that students encounter in everyday life, e.g., fare dodging on a train, illegal downloading, minor thefts, student loan fraud. In each vignette, one of the five dimensions was manipulated, whereas the other dimensions were held constant. This resulted in two versions of the same vignette that formed a vignette pair; see Table 1 for an example.

Each vignette contained additional information suggesting a certain mundane behavior, which was independent of the criminal behavior depicted in the same vignette. These behaviors were not against the law, but problematic for other reasons, e.g., because they required self-control (examples: not eating cake on a diet, starting an unpleasant conversation, showing courage, not cheating on someone). These mundane behavior options were added to ensure that the RTs could be assessed validly, as participants were not able to anticipate what they would be asked

[2] Preregistration Study 1: https://osf.io/w6c8a; Study 2: https://osf.io/jg6wk. Note: In the preregistration, the dimension “incentive” was provisionally named “impulse”.

about – either the criminal behavior option or the mundane behavior option[3]. Another advantage of adding the mundane behavior information was that participants did not immediately recognize that the study was about delinquency, which may have reduced response bias.

**Table 1**

*Example of a vignette pair in which the dimension* Level of Punishment *was manipulated*

| Vignette version: Level of Punishment | | |
|---|---|---|
| Low | High | Subsequent question |
| *You're in a relationship,* but now you're in the car *on your way to meet your affair in another city. Your best friend advised you to finally talk to your partner about the affair.* While you are thinking about it, you reach a section on the motorway where you are only allowed to drive 80 km/h for anti-noise reasons. You have **no demerit points** in the central register. *You assume that your partner has already suspected that you are having an affair.* | *You're in a relationship*, but now you're in the car *on your way to meet your affair in another city. Your best friend advised you to finally talk to your partner about the affair.* While you are thinking about it, you reach a section on the motorway where you are only allowed to drive 80 km/h for anti-noise reasons. You have **five demerit points** in the central register. *You assume that your partner has already suspected that you are having an affair.* | Do you drive too fast? |

*Note*. The manipulated element is in bold, the distracting information suggesting a mundane behavior option is in italics.

We constructed four pairs of vignettes per dimension, which made a total of 20 vignette pairs. One vignette version had a high value on the manipulated dimension and the other one had a low value on the same dimension, while values on all other dimensions were identical. It was not possible to manipulate all dimensions within one vignette, because as a result, the

[3] As a result, it was unclear to the participants which of the information was relevant in each vignette. Our aim was to postpone the appraisal processes related to decision making until the question was presented. Otherwise, the RTs would have been influenced by the position of the appraisal information within the vignettes.

vignettes would no longer have been realistic in content. The vignette versions with high values on the dimensions *Incentive*, *Legality* and *Morality* and low values on the dimensions *Likelihood* and *Level of Punishment* should favor behavior execution, whereas vignettes with opposite values should hinder behavior execution. Each participant was presented only one version from each pair. Per dimension each participant was presented two vignettes with low values and two vignettes with high values on the crucial appraisal dimension.

After the presentation of a manipulated vignette, participants were asked whether or not they would execute the criminal behavior option (e.g., 'Do you keep the money?'). In the manipulated vignette pairs, the mundane behavior options were not queried. In addition to the 20 manipulated vignette pairs, 20 single distractor vignettes, which were designed equivalently (i.e., contained information alluding to criminal as well as mundane behavior), were presented. The distractor vignettes were not manipulated, so the same version of the vignette was presented to all participants. For the distractor vignettes, participants were asked whether or not they would execute the mundane behavior. This means that in 50% of cases each participant was asked about the criminal behavior and in 50% of cases about the mundane behavior, which was irrelevant to our study.

**Manipulation Check Study.** In order to check whether the two versions of a vignette pair did in fact differ regarding the dimension we intended to manipulate, we asked an independent sample of $N = 17$ students to rate all vignette versions regarding their values on the five relevant appraisal dimensions using a five-point Likert scale. We expected both vignette versions of each pair to differ on the one dimension we intended to manipulate, while there should be no difference on most other dimensions. However, some of the dimensions are naturally confounded with each other and cannot be manipulated orthogonally (e.g., *Legality* and *Level of Punishment*).

Repeated measures ANOVAs with within-participant contrasts showed that 15 of the 20 pairs of vignettes differed significantly in the dimension we intended to manipulate (p-values $< .05$). However, for 11 of these 15 vignette pairs there was also a significant change in at least one other dimension that was not manipulated (see Table C1 in Appendix C for an

overview). We excluded the five vignettes for which the manipulation had failed from all analyses in which the dimension-specific decision behavior was considered.[4]

**Dependent Variable.** The dependent variable in the main study was the frequency of positive responses to the question of behavior execution in the manipulated vignette pairs. The response categories "No" and "Yes" ("Ja" and "Nein" in German) were presented in boxes next to each other on a computer screen – "No" to the left, "Yes" to the right. The response was made via mouse click. It was also possible to take part in the study on a tablet pc or smartphone. In this case the presentation was the same, but the response was given by touching the answer boxes on the display with a finger. Responses to the distractor vignettes were not analyzed.

**Other Measures.** We measured the Big Five personality dimensions with the German short version of the Big Five Inventory (BFI-K; Rammstedt & John, 2005), which comprises 21 items. Respondents indicated the degree to which each item applied to them along a Likert scale ranging from 1 = *very untrue* to 5 = *very true*. The BFI-K consists of five scales: *openness* (five items, α = .77), *conscientiousness* (four items, α = .66), *extraversion* (four items, α = .82), *agreeableness* (four items, α = .70), *neuroticism* (four items, α = .77).

Psychopathy was assessed with the 40-item German short version of the Psychopathic Personality Inventory–Revised (PPI-R-40; Eisenbarth et al., 2015; α = .79) with the three higher-order dimensions *Fearless Dominance* (13 items, α = .75) and *Self-Centered Impulsivity* (22 items, α = .81) and *Coldheartedness* (five items, α = .70). Respondents indicated the degree to which each item applied to them along a four-point Likert scale ranging from 1 = *false* to 4 = *true*.

We also assessed demographic information in terms of gender and age. For information on delinquent lifestyles we asked for the number of previous convictions and drug consumption. Therefore, we asked in how many days during the last 30 days the participants had used the following drugs: alcohol, tobacco, cannabis, and illegal hard drugs such as crystal meth or

[4] We also performed these analyses with all 20 vignettes. The results were very similar to the results reported here and did not lead to different results in testing research hypotheses.

cocaine. Furthermore, participants indicated if they were currently students or had studied within the last two years.

### *Procedure*

At the beginning, participants were informed about the conditions of the study (estimated completion time, information that participation was voluntary and anonymous) and gave their consent. They were told the study was on "decision making in different situations". After that, they were asked to indicate their age, gender, and student status. If participants stated that they were neither students nor had they studied within the last two years, they were told that they could not participate and directed to the end of the study. The remaining participants were instructed to carefully read the subsequent scenarios and told that there were no right or wrong answers. They were instructed to try to respond in the way they would decide if the situation was real. In every trial, the description of the scenario (vignette) was presented first. When the participant clicked on the "continue" box, the vignette disappeared and the crucial question with the answer categories "No" and "Yes" appeared.

The 40 vignettes from the 20 manipulated vignette pairs were divided into two sets of vignettes to make sure that only one version of a vignette was assigned to each participant. Both the allocation of the sets to the participants and the order of the vignettes within a set were randomized. After participants had completed the vignette task, the questionnaire materials were displayed in the following order: BFI-K, PPI-R-40, questions about previous convictions and drug consumption.

## Method Study 2

### *Participants*

After the exclusion of two participants who had stated they had not participated seriously, the data of $N = 199$ participants (80% female, 20% male) were included in the analyses. Participants were aged between 18 and 48 years ($M = 24.08$, $SD = 3.71$) and were currently students or had studied within the past two years. A total of 81% of participants reported having consumed alcohol at least once in the last 30 days (tobacco: 24%, cannabis: 16%, hard drugs: 3%). Three participants stated that they had at least one conviction. Participants had the

opportunity to win an ice cream machine worth €50 or one of two Amazon vouchers worth €25. Psychology students received partial course credit.

### *Materials, Design and Procedure*

We used the same materials and procedure as in Study 1 except for the following differences: The vignette texts and questions were recorded on tape and presented acoustically instead of in written text form as in Study 1. All vignettes were recorded by the same female speaker. At the beginning of the study, the participants were informed that the study contained audio files, so they could make sure that headphones or speakers were available. All other materials (instructions, questionnaires) were presented in form of written text and in the same order as in Study 1. The only addition was a seriousness check consisting of one question presented at the very end of the study (Aust et al., 2013). This question allowed the participants to indicate whether they had taken part seriously or whether they had just clicked through and would rather we not use their data. Participants were informed that their response to the seriousness check item would not have any negative consequences (e.g., exclusion from the lottery).

The internal consistencies of the BFI-K (Big Five) scales ranged from α = .71 (agreeableness, conscientiousness) to .85 (extraversion). The overall Cronbach's α for the PPI-40-R (psychopathy) was .81, and the Cronbach's α of one of the subscales was somewhat lower (Fearless Dominance: α = .78, Self-Centered Impulsivity: α = .80, Coldheartedness: α = .61).

## Results

### *Dimensions' Influence on Decision Making*

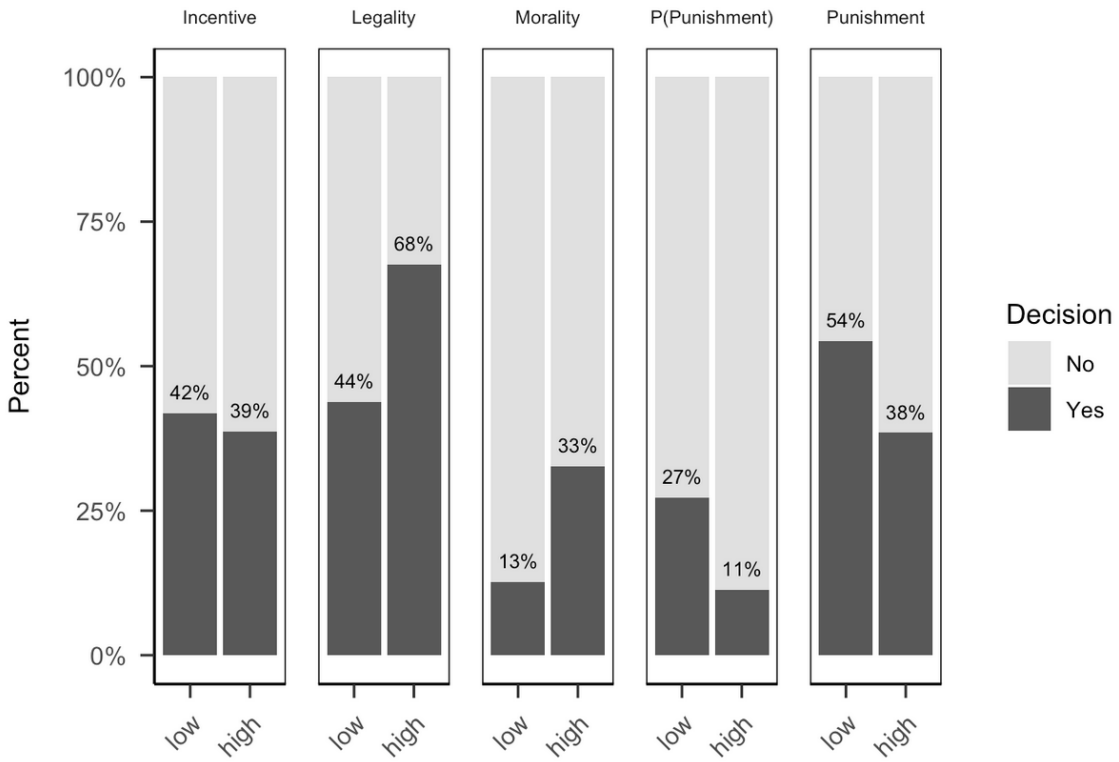
The data from Studies 1 and 2 were aggregated into a combined dataset of $N = 299$ cases. To obtain an overview of participants' decision making, we first examined the answers independently of the dimensions. Across all vignettes and all participants there were more negative ($k = 3{,}627$, 65%) than positive ($k = 1{,}996$, 35%) responses. Our participants therefore showed a tendency to reject rather than accept the criminogenic opportunities depicted in the vignettes.

Next, it was analyzed whether the manipulations had an impact on the frequency of criminal decisions. The values of the dimensions *Likelihood of Punishment* and *Level of Punishment* were recoded so that according to our hypothesis high values (i.e., a low likelihood and a low

level of punishment) are expected to favor the execution of criminal behavior. Across all dimensions, in vignette versions with low values on the crucial appraisal dimension, participants decided less often in favor of the criminal behavior option than in versions with high values ($\chi^2(1, 5{,}623) = 133.08$, $p < .001$, $\phi = -0.15$).

**Figure 3**

Percentage of positive (i.e., criminal) decisions by dimension.




Note. Low values on the dimensions *Incentive*, *Legality* and *Morality* and high values on the dimensions *Likelihood of Punishment* (P(Punishment)) and *Level of Punishment* (Punishment) were expected to lead to a lower percentage of positive responses.

The frequencies of positive (i.e., criminal) and negative decisions by dimension are shown in Figure 3. To check whether participants' decisions were influenced by the five manipulated dimensions, we performed $\chi^2$-tests for each dimension. They revealed that the manipulation led to significant effects on decisions in the predicted directions for the dimensions *Legality* ($\chi^2(1, 1{,}114) = 63.18$, $p < .001$, $\phi = -0.24$), *Morality* ($\chi^2$ (1, 848) = 55.37, $p < .001$, $\phi = -0.26$), *Likelihood of Punishment* ($\chi^2$ (1, 850) = 33.92, $p < .001$, $\phi = -0.20$) and *Level of Punishment* ($\chi^2$ (1, 841) = 20.49, $p < .001$, $\phi = -0.16$). However, there was no significant difference for

*Incentive* ($\chi^2$ (1, 564) = 0.47, $p$ = .491, $\phi$ = 0.03). Thus, Hypothesis 1 was confirmed for four of the five appraisal dimensions.

### *Associations between Person Factors and Decision Making*

The bivariate Spearman correlations between criminal decision making (frequency of positive decisions) and person factors (psychopathy, Big Five, demographic and lifestyle variables) are presented in Table 2. As predicted, participants with higher psychopathy scores showed a higher frequency of positive (i.e., criminal) decisions than participants lower in psychopathy (Hypothesis 2; $r$ = .39, $p$ < .001). Of the three subscales that constitute the total psychopathy score, *Self-Centered Impulsivity* was most strongly correlated with criminal decisions ($r$ = .37, $p$ < .001). There was also a small but significant negative association between criminal decision making and *Agreeableness* ($r$ = -.13, $p$ = .026). However, we did not find significant associations with other Big Five personality traits nor with gender and age (all $p$s > .05). The lack of correlations with demographic variables could be due to limited variance in our sample: Participants were mean aged 24 years with $SD$ = 3.53 and 80% were female. Furthermore, participants who had used drugs more frequently within the last 30 days showed a higher frequency of criminal decisions than participants who used less drugs ($r$ = .29, $p$ < .001). The correlation between previous convictions and criminal decision making ($r$ = .14, $p$ = .018) was driven by five participants who had at least one conviction.

**Table 2**

*Spearman correlations of all measures from Studies 1 and 2.*

| | *M* | *SD* | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. | 13. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Criminal Decisions | 1.36 | 0.15 | -- | | | | | | | | | | | | |
| **Psychopathy** | | | | | | | | | | | | | | | |
| 2. Psychopathy (total) | 2.12 | 0.29 | .39** | -- | | | | | | | | | | | |
| 3. Fearless Dominance | 2.45 | 0.47 | .17** | .66** | -- | | | | | | | | | | |
| 4. Self-Centered Impulsivity | 1.97 | 0.38 | .37** | .79** | .15** | -- | | | | | | | | | |
| 5. Coldheartedness | 1.89 | 0.52 | .19** | .35** | .14* | .09 | -- | | | | | | | | |
| **Big Five** | | | | | | | | | | | | | | | |
| 6. Openness | 3.94 | 0.71 | -.02 | .04 | -.04 | .13* | -.16** | -- | | | | | | | |
| 7. Conscientiousness | 3.65 | 0.66 | -.09 | -.24** | -.06 | -.37** | -.08 | .02 | -- | | | | | | |
| 8. Extraversion | 3.41 | 0.88 | .10 | .25** | .47** | .01 | -.01 | .07 | .11 | -- | | | | | |
| 9. Agreeableness | 3.05 | 0.84 | -.13* | -.19** | -.17** | -.33** | -.16** | .16** | -.09 | .32** | -- | | | | |
| 10. Neuroticism | 3.19 | 0.92 | .01 | -.30** | -.68** | .12* | -.15* | .02 | -.14* | -.40** | -.30** | -- | | | |
| **Demographics** | | | | | | | | | | | | | | | |
| 11. Gender | -- | -- | -.11 | -.26** | -.27** | -.11 | -.12* | .12* | .23** | .03 | .03 | .20** | -- | | |
| 12. Age | 24.01 | 3.53 | -.01 | .03 | -.01 | .04 | .08 | .05 | -.04 | .02 | -.02 | -.05 | -.02 | -- | |
| **Lifestyle** | | | | | | | | | | | | | | | |
| 13. Drugs (total) | 2.48 | 3.54 | .29** | .34** | .22** | .30** | .04 | -.03 | -.20** | .25** | -.08 | -.11 | -.23** | .04 | -- |
| 14. Convictions | 0.02 | 0.19 | .14* | .14* | .01 | .15** | .07 | -.04 | -.05 | .02 | -.14* | -.01 | .00 | .07 | .12* |

*Note.* 1. Criminal Decisions: Mean decision across all vignettes (1=No, 2 =Yes); 11. Gender: 1=male, 2= female; 13. Drugs (total)= mean substance use within the last 30 days (alcohol, tobacco, cannabis and other drugs); if not stated otherwise: higher values indicate higher manifestations; $N = 299$; * $p < .05$, ** $p < .01$.

### *Interactions between Person Factors and Decision Making*

In order to examine Hypothesis 3 stating that participants with a high psychopathy score react in a less sensitive way to a manipulation of the dimensions, we conducted separate analyses for participants with high and low psychopathy scores and compared them to each other. Participants whose scores were in the first quartile of the distribution were defined as low scorers (≤ 1.92, $n = 83$), and participants whose scores were in the fourth quartile were defined as high scorers (≥ 2.30, $n = 79$).

The results of the $\chi^2$-tests, which examined across all dimensions and separately for each dimension whether low and high scorers differed regarding their sensitivity towards the manipulation, are displayed in Table 3.

**Table 3**

*Dimensions' influence on decision making of participants with low and high psychopathy scores: χ2-tests*

| Dimension | Psychopathy | *k* | $\chi^2$ | *p* | ϕ | Δϕ |
|---|---|---|---|---|---|---|
| Across all dimensions | low | 1570 | 48.54 | .00*** | -.18 | -0.08* |
| | high | 1479 | 14.91 | .00*** | -.10 | |
| Incentive | low | 155 | 0.58 | .45 | -.08 | -0.23* |
| | high | 150 | 2.68 | .10 | .15 | |
| Legality | low | 315 | 30.06 | .00*** | -.32 | -0.25** |
| | high | 294 | 1.29 | .26 | -.07 | |
| Morality | low | 237 | 9.11 | .00** | -.21 | 0.11 |
| | high | 220 | 21.18 | .00*** | -.32 | |
| Likelihood of Punishment | low | 237 | 14.64 | .00*** | -.26 | -0.05 |
| | high | 227 | 8.89 | .00** | -.21 | |
| Level of Punishment | low | 230 | 8.47 | .00** | -.20 | -0.14 |
| | high | 222 | 0.62 | .43 | -.06 | |

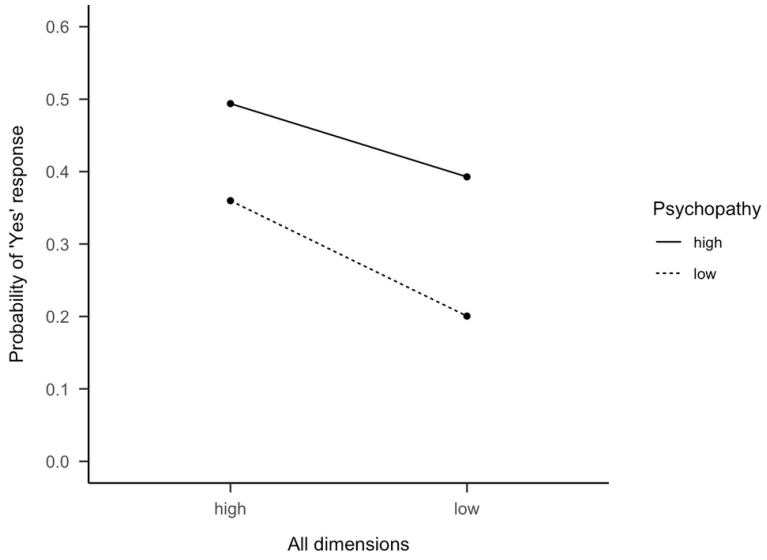
Note. low/high psychopathy = scores in the first quartile (≤ 1.92) and in the fourth quartile (≥ 2.30) of the distribution; *k*= number of decisions included; the five vignettes for which the manipulation check had failed were only included in the $\chi^2$-tests across all dimensions. * $p < .05$, ** $p < .01$., *** $p < .001$

Across all vignettes, participants with low psychopathy scores were more strongly influenced by the manipulation of the dimensions than participants high in psychopathy (Δϕ = -0.08, $z = -2.25$, $p = .024$, two-tailed; tested using *cocor* R package, which transforms coefficients to

Fisher's *z*-scores and assesses differences between them; Diedenhofen & Musch, 2015). This interaction effect is displayed in Figure 4. Accordingly, Hypothesis 3 was confirmed.

**Figure 4**

*Probability of a "Yes" response (i.e., criminal decision) for participants high and low in psychopathy across all dimensions*




Separate analyses for all five dimensions showed that the difference between both groups was highest for *Legality* ($\Delta\phi$ = -0.25, $z$ = -3.21, $p$ = .001). Participants low in psychopathy were significantly influenced by the manipulation of the dimension *Legality*, whereas participants with high psychopathy scores were not (Figure 5). The difference of $\phi$-values was also significant for *Incentive* ($\Delta\phi$ = -0.23, $z$ = -2.00, $p$ = .046). Interestingly, it seems that when incentives were low, participants high in psychopathy tended to make more criminal decisions, whereas participants low in psychopathy showed the opposite tendency (Figure 6). However, separate $\chi^2$-tests for both high and low scorers were not significant. Furthermore, participants low in psychopathy made significantly fewer criminal decisions when there was a high punishment, whereas participants high in psychopathy were not significantly deterred by a high punishment. However, the difference of $\phi$-values was not significant ($\Delta\phi$ = 0.14, $z$ = -1.51, $p$ = .132).

**Figure 5**

*Probability of a "Yes" response for participants high and low in psychopathy for vignettes in which Legality was manipulated*

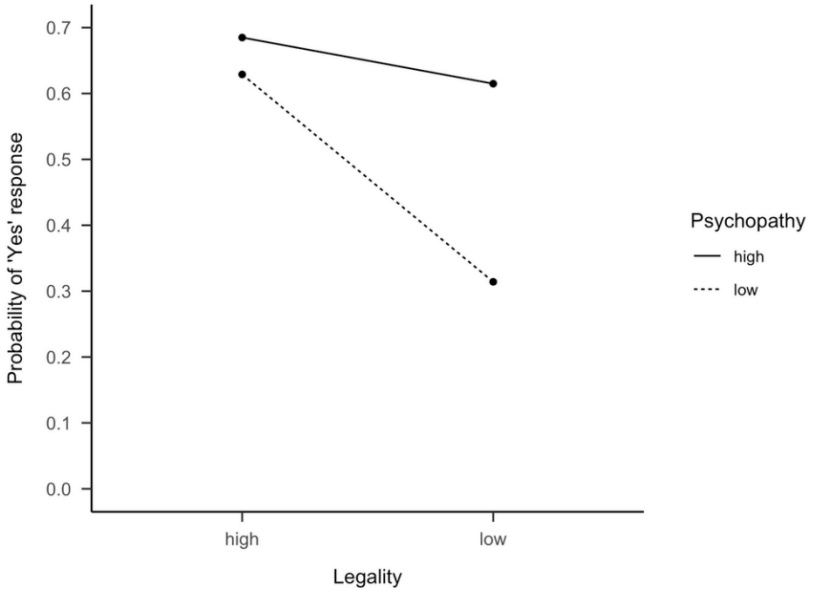
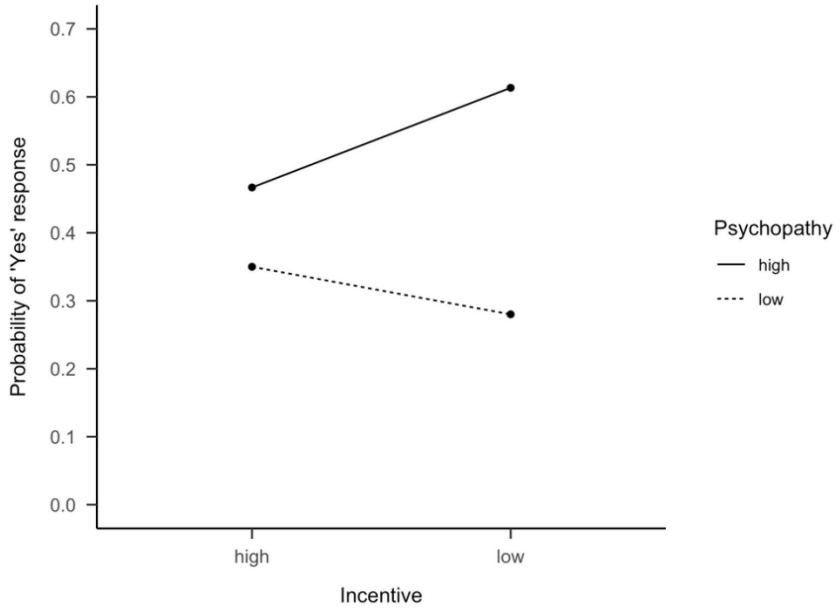



**Figure 6**

*Probability of a "Yes" response (i.e., criminal decision) for participants high and low in psychopathy for vignettes in which Incentive was manipulated*



**Discussion**

In Studies 1 and 2, we found that four out of five of the appraisal model's dimensions in fact had an impact on everyday criminal decision making in students. The effect size was largest for the dimension *Morality*. However, there was no influence of the dimension *Incentive*, and, descriptively, low incentives even led to more criminal decisions than high incentives. Exploratory analyses revealed that this tendency was mainly driven by participants high in psychopathy. Additionally, it should be noted that due to failed manipulation checks two of the four vignettes in which the dimension *Incentive* should have been manipulated had to be excluded from the analyses. Thus, both the variety of situations and the statistical power were smaller in *Incentive* than in the other dimensions, where we used three or four vignettes per dimension.

Psychopathy (especially *Self-Centered Impulsivity*) was positively correlated with the tendency to decide in favor of the criminal action depicted in the vignettes, but there were no or only minor correlations with Big Five personality traits, gender, and age. Both participants with high and low psychopathy scores were influenced by the manipulation of the dimensions in the predicted direction, but this impact was stronger for participants with low psychopathy scores. This difference between high and low scorers regarding sensitivity for a manipulation of the dimension was most evident for the dimension *Legality*. Whether an action was illegal or legal was very relevant to people low in psychopathy but did not make a significant difference to participants high in psychopathy, who generally showed a higher base rate of criminal decisions. Furthermore, if incentives were low, participants high in psychopathy even tended to make more criminal decisions, whereas participants low in psychopathy showed the opposite tendency (the differences between high and low scorers was significant; however, the single paths were not). Further research should replicate this finding with a larger number of vignettes and explore whether other constructs such as impulsivity and sensation-seeking play a role in this context (because to people who score high in psychopathy, carrying out a criminal action that brings little benefit might actually be particularly appealing, simply for the sake of doing something illegal).

There were no interactions between psychopathy and the dimensions *Morality* and *Likelihood of Punishment*. Participants low in psychopathy were influenced by the *Level of Punishment*, whereas participants high in psychopathy were not. This finding is consistent with literature

on punishment insensitivity of people scoring high in psychopathy. The inability to learn from punishment has been recognized by Cleckley (1988) as a central feature of psychopathy, has been confirmed by experimental studies using learning tasks (e.g., Blair et al., 2004; Blair et al., 2006; Newman & Kosson, 1986), and linked to structural and functional impairments in the prefrontal cortex (Umbach et al., 2015).

After using the vignette paradigm to investigate our model assumptions in the area of general delinquency and obtaining promising results regarding the impact of dimensions and their interaction with psychopathy, we transferred it to the domain of dissexual behavior in Study 3.

## Study 3

Not every form of sexually offensive behavior is prosecuted or subject to prosecution. Nevertheless, the recent #MeToo movement shows that many people have experienced the disregard of their well-being or sexual self-determination through the sexual acts of another. This sexual expression of a failure to conform to social norms is called dissexuality (Beier, 1998). The term includes sexual offences but is not limited to them.

As dissexual behavior occurs frequently in everyday situations and usually arises from situational opportunities, we assume that the appraisal theory can be applied to this domain of dissocial behavior as well. According to our framework model, appraisals can be influenced not only by stable person factors (such as psychopathy or self-control) but also by dynamic person factors (such as arousal). This study offers the opportunity to investigate the effect of sexual arousal on dissexual decision making.

Ariely and Loewenstein (2006) showed that sexually aroused men are attracted to a wider range of sexual stimuli and activities, and are more willing to engage in unsafe sex and in morally questionable behavior (e.g., expressing love to a woman to increase the chance of having sex with her). This disinhibitory effect was replicated by Imhoff and Schmidt (2014), who argued that situational sexual arousal may function as a previously ignored risk factor for socially inadequate, unhealthy, and manipulative sexual behavior. The mechanism underlying this effect could be that sexual arousal increases perceptions of sexual pleasure as a currently important benefit. This was found by Bouffard (2002), who conducted a study in which participants first viewed sexually arousing (or control) stimuli, then read a date-rape scenario and estimated their likelihood of engaging in several sexually coercive tactics, and finally

listed potential costs and benefits and rated their certainty, severity, and importance. Sexual arousal increased participants' likelihood of engaging in coercive behavior and perceptions of sexual pleasure as a benefit. In a similar study, the effect that sexual arousal altered the perception of sexual pleasure as an important benefit was replicated, though arousal did not seem to have any impact on perceived costs (Bouffard, 2011). However, the results of the few studies on mediating or moderating effects of sexual arousal are rather inconsistent: Loewenstein et al. (1997) found that neither the perception of costs nor benefits mediated the relationship between arousal and sexual aggression in another date-rape scenario. It is therefore important to investigate this more closely.

People are usually not aware of the impact that their own sexual arousal has on their decisions and, therefore, they are unable to counteract it (Ariely & Loewenstein, 2006; Loewenstein et al., 1997). This "hot-cold empathy gap", i.e., the phenomenon that if someone is in a "cool", unemotional state, he or she has difficulties predicting his/her own behavior when in a "hot", emotional state, applies to several areas of life. When we are not hungry, afraid, or angry, we have trouble imagining how we would feel or act when we experience these states (Loewenstein et al., 1997). In the current study, we examined whether the appraisal dimensions still affect decisions in a field in which the heat of the moment is crucial. We decided to not intentionally induce sexual arousal, as was done in some studies (e.g., using erotic imagery, Loewenstein et al., 1997, or narratives, Imhoff & Schmidt, 2014), because our participants would inevitably be exposed to potentially sexually arousing content during the task. Therefore, we did not use sexual arousal as an experimental variable but as a dynamic (state) and differential (trait) person factor that was expected to predict dissexual decisions. As further potential predictors for dissexual behavior we assessed participants' sex drive and sexual disgust sensitivity. Sexual disgust is generally reduced under circumstances of sexual arousal (De Jong et al., 2013; Stevenson et al., 2011), and negatively correlated with sex drive (Eickmeier et al., 2019).

In addition, we decided to measure morality and self-control as stable person factors that should be negatively related to dissexual decision making. We chose these variables because they play an important role in Wikström's SAT. According to SAT, self-control (in addition to external controls) is only relevant if one does not have any moral concerns regarding a potential action (Wikström & Treiber, 2007). Empirical research has largely confirmed that

personal morals are more fundamental than self-control abilities in the explanation of criminal conduct, and also found evidence for the interaction between morality and self-control (De Li, 2004; Hirtenlehner & Kunz, 2016; Schoepfer & Piquero, 2006; Svensson et al., 2010; Tittle et al., 2010; Wikström & Svensson, 2010). Moreover, the importance of self-control as a predictor of crime was confirmed in two meta-analyses (Pratt & Cullen, 2000; Vazsonyi et al., 2017). Because it is still unclear through which mechanisms self-control is linked to delinquency, we now investigate the causal role of the appraisal of dissexual opportunities. It is conceivable that people with low self-control put more weight on incentives when they make their decision, or they could appraise even small incentives as very high. If this is the case, the manipulation of the dimensions would have different effects on decisions of people high or low in self-control. We also explored whether people with high moral standards are insensitive to a manipulation of costs and benefits, which would be in line with SAT. As findings on morality and self-control are mainly based on samples of adolescents and young adults, we aimed to include men of all age groups to obtain a more comprehensive picture. In addition, SAT research has largely neglected sexual delinquency as a type of offence. We are therefore closing this gap with our study. Our sample is only male because the vast majority of sexual offences are perpetrated by men (Oliver, 2007). Our hypotheses, which were preregistered (see Appendix),[5] are as follows:

Hypothesis 1: A manipulation of each of the five dimensions of the appraisal model (*Incentive*, *Legality*, *Morality*, *Likelihood*, and *Level of Punishment*) within the vignettes will lead to a change of the frequency of positive decisions toward dissexual behavior.

Hypothesis 2: Participants with high morality, self-control and sexual disgust scores will show a lower frequency of dissexual decisions than participants with low scores (i.e., a negative correlation). Sex drive is expected to be positively correlated with dissexual decisions.

Hypothesis 3: Sexual arousal experienced during the study will be positively correlated with dissexual decisions. Since we assume that sexual arousal will continuously increase during the task, vignettes presented later are expected be related to more dissexual behavior responses.

[5] Preregistration Study 3: https://osf.io/n783y/?view_only=0c1f5f2edf76466195530b3ea635215d

Therefore, the individual positions of the vignettes in the sequence should predict the frequency of dissexual decisions.

Moreover, we investigated whether participants high in morality and self-control were influenced differently than participants low in morality and self-control by a manipulation of the dimensions. We expected in particular:

Hypothesis 4 a): Due to a low base rate of endorsing dissexual behavior (and in line with SAT), participants with high morality scores will show no or only weak responses to a manipulation of the three dimensions *Incentive*, *Likelihood*, and *Level of Punishment* in comparison to participants low in morality.

Hypothesis 4 b): Participants with a high self-control score will show no or only weak responses to a manipulation of the dimension *Incentive* in comparison to participants low in self-control.[6]

## Method

### *Participants*

From a total of 742 participants, 685 men aged 18 to 81 ($M = 37.33$, $SD = 10.87$)[7] who were sexually interested in women were in the final sample. We only included heterosexual or bisexual men who had participated seriously in the study and – as an indicator for sex drive – who stated they had at least one orgasm per week on average in the last six months (including masturbation). A total of 57 participants did not meet these inclusion criteria and were excluded from further analyses.

In this final sample, the majority of participants exclusively had heterosexual contacts (60%) and fantasies (56%). Nearly one-third (32%) stated they occasionally had homosexual contacts, and 33% occasionally had homosexual fantasies. Only 4% stated they had

---

[6] Hypotheses 4a) and 4b) contain the same typing error in the preregistration ("Due to a low base rate (…), participants with a high morality (self-control) score should show no or only weak responses (…) in comparison to participants *high* in morality"). Of course, we intended to compare participants with high levels of morality (self-control) to participants with *low* levels.

[7] Information on age and educational level was missing for 28 participants.

homosexual contacts more than occasionally (fantasies: 7%). Only 3% were equally hetero- and homosexual in terms of behavior (fantasies: 5%). People from all German educational classes were represented in the sample (no degree: < 1%, "Hauptschule" / basic school degree: 9 %, "Realschule" / secondary school degree: 29 %, "Abitur" / university entrance qualification: 28 %, Bachelor's degree: 15%, Master's degree: 14%). Participants were recruited via Facebook, personal contacts, and a casual online dating website's newsletter. Users of this platform can arrange to have sex with other users. The participants were offered the opportunity to win a portable charcoal grill worth €40 or one of two €30 Amazon vouchers.

### *Materials and Design*

**Vignettes.** The vignettes were designed according to the same principles as the vignettes used in the first two studies, only that the content and subsequent questions dealt with sexual delinquency and dissexual behavior instead of general delinquency (see Table 4 for an example of a vignette). The questions were, e.g., about whether participants would make a sexist remark, send sexually suggestive text messages, touch someone without her consent, urge someone to engage in certain sex practices, continue sex although she wants to stop, or buy someone drinks so she would agree to have sex.

As in Studies 1 and 2, as a manipulation check an independent sample of $N = 23$ men rated the vignettes regarding their values on the five appraisal dimensions using a five-point Likert scale. Repeated measures ANOVAs with within-participant contrasts showed that 17 of the 20 pairs of vignettes significantly differed in the dimension we intended to manipulate ($p$-values < .05). For nine of these 17 vignette pairs there was also a significant change in at least one other dimension which was not manipulated (see Table C2 in Appendix C). We excluded the three vignettes for which the manipulation had failed from all analyses in which the dimension-specific decision behavior was considered.[8]

[8] We additionally performed these analyses with all 20 vignettes. The results were very similar to the results reported here and did not lead to different results in testing research hypotheses.

**Table 4**

*Example of a vignette pair in which the dimension* Likelihood of Punishment *was manipulated.*

| Vignette version: Likelihood of Punishment | | |
|---|---|---|
| Low | High | Subsequent question |
| Today you took a day off and went to the nearby beach. There you enjoy the beautiful summer day. There is not much going on and apart from you there are only two other women. One of the women is very attractive and is sunbathing naked. You would like to take a photo to masturbate with it later. **Both the naked woman and her companion seem to have fallen asleep.** *As the beach fills up a bit in the afternoon, you suddenly discover a work colleague who has been calling in sick for weeks. You have to cover all his duties and you are annoyed because your colleague doesn't look ill at all.* | Today you took a day off and went to the nearby beach. There you enjoy the beautiful summer day. There is not much going on and apart from you there are only two other women. One of the women is very attractive and is sunbathing naked. You would like to take a photo to masturbate with it later. **The naked woman is asleep, but her companion is awake and looking at the water.** *As the beach fills up a bit in the afternoon, you suddenly discover a work colleague who has been calling in sick for weeks. You have to cover all his duties and you are annoyed because your colleague doesn't look ill at all.* | Do you take a photo? |

*Note*. The manipulated element is in bold, the distracting information suggesting a mundane behavior option is in italics.

**Dependent Variable.** Again, we measured the frequency of positive responses to the question of behavior execution in the manipulated vignette pairs.

**Other measures.**

***Sexual Arousal.*** We assumed that participants would become sexually aroused during the vignette task. To assess this effect, we generated two items and asked participants to rate their sexual arousal on a five-point Likert scale (e.g., "I am sexually aroused right now" $\alpha = .75$). The participants also indicated their non-sexual arousal (e.g., "I am excited right now") and their affective state (e.g., "I feel good right now"), also measured by two items each. For all items we used a five-point Likert scale (1 = *untrue*; 5 = *very true*).

***Sexual Orientation***. Sexual orientation was measured using the two-item Kinsey Sexual Ratings scale (Kinsey et al., 1948), which asks about ideas and fantasies in relation to the male and/or female sex, and about actual sexual contacts with the male and/or female sex. The response scale represents a continuum of sexual orientation ranging from 1 = exclusively heterosexual to 7 = exclusively homosexual. Additionally, participants had the option to indicate that they had no sexual contacts or reactions.

***Sociosexual orientation.*** The Revised Sociosexual Orientation Inventory (SOI-R; Penke & Asendorpf, 2008) was used to assesses three facets of sociosexuality using three items per facet: past behavior in terms of number of casual and changing sex partners (Subscale *Behavior*, e.g., "With how many different partners have you had sex within the past 12 months?", $\alpha = .81$), explicit attitude towards uncommitted sex (Subscale *Attitude,* e.g., "Sex without love is OK", $\alpha = .73$), and sexual desire for people without having a romantic relationship (Subscale *Desire,* e.g., "In everyday life, how often do you have spontaneous fantasies about having sex with someone you have just met?", $\alpha = .82$) (Penke, 2011). For all items, a five-point response scale was used. Cronbach's alpha for the total score was $\alpha = .81$.

***Sex drive***. As an indicator for sex drive we measured participants' total sexual outlet (TSO; Kafka, 1997). With an open response category, participants were asked to indicate their average number of orgasms per week experienced in the last six months including masturbation (TSO/week) and the maximum number of orgasms per week experienced since age 15 including masturbation (TSO/max). In order to robustly identify outliers, the median absolute deviation was calculated (Leys et al., 2013), resulting in cut-offs for outliers of $\geq 13$ orgasms for TSO/week and $\geq 37$ orgasms for TSO/max. Furthermore, TSO-values were excluded due to implausibility if TSO/week was higher than TSO/max.[9]

[9] A total of 164 outliers (24%) were excluded for TSO/week and 151 outliers (22%) were excluded for TSO/max. It appeared to us that several participants accidentally had not indicated the weekly number of orgasms, but the cumulated number of orgasms during the last six months or since age 15.

***Sexual disgust.*** We used the sexual disgust subscale from the Five-Factor Disgust Scale (5-FES; Fünf-Faktoren Ekelskala; Eickmeier et al., 2019). Participants were asked to indicate how disgusting they considered six different behaviors (e.g., "Using sex toys"; α = .67) on a Likert scale from 1 = *not disgusting at all* to 5 = *extremely disgusting*.

***Self-control.*** To assess participants' dispositional self-control capacities we used the German short version of the Self-Control Scale by Bertrams and Dickhäuser (2009, SCS-K-D; English version by Tangney et al., 2004). The scale consists of 13 items (e.g., "I am lazy"; α = .78) and responses were made on a five-point Likert scale (1 = *not at all*; 5 = *very much*).

***Morality.*** We used the morality scale by Hirtenlehner and Kunz (2016), whereby we modified the original ten-point Likert scale to a five-point Likert scale. Participants' moral values were assessed by asking them how wrong they think it is to commit seven specific misconducts (1 = *not at all wrong*; 5 = *very wrong*) which varied in the degree of seriousness from "fraudulently obtaining social benefits for which one is not eligible" to "beating and physically injuring other persons" (α = .88).

***Seriousness Check.*** As in Study 2, we used the seriousness check item (Aust et al., 2013) to determine if all participants had taken part seriously or if they e.g., just clicked through to read the vignettes.

### *Procedure*

At the beginning, participants gave their informed consent to take part in the study. They were informed that the study was about men's decision making in social situations and that the study also contained sexual content. After participants had provided information on their gender, age, and educational level, they were presented the vignettes and asked if they would carry out the respective behavior or not. Immediately after participants had completed the vignette part of the study, they were asked to indicate their arousal. All measures were implemented in the same order in which they are described above.

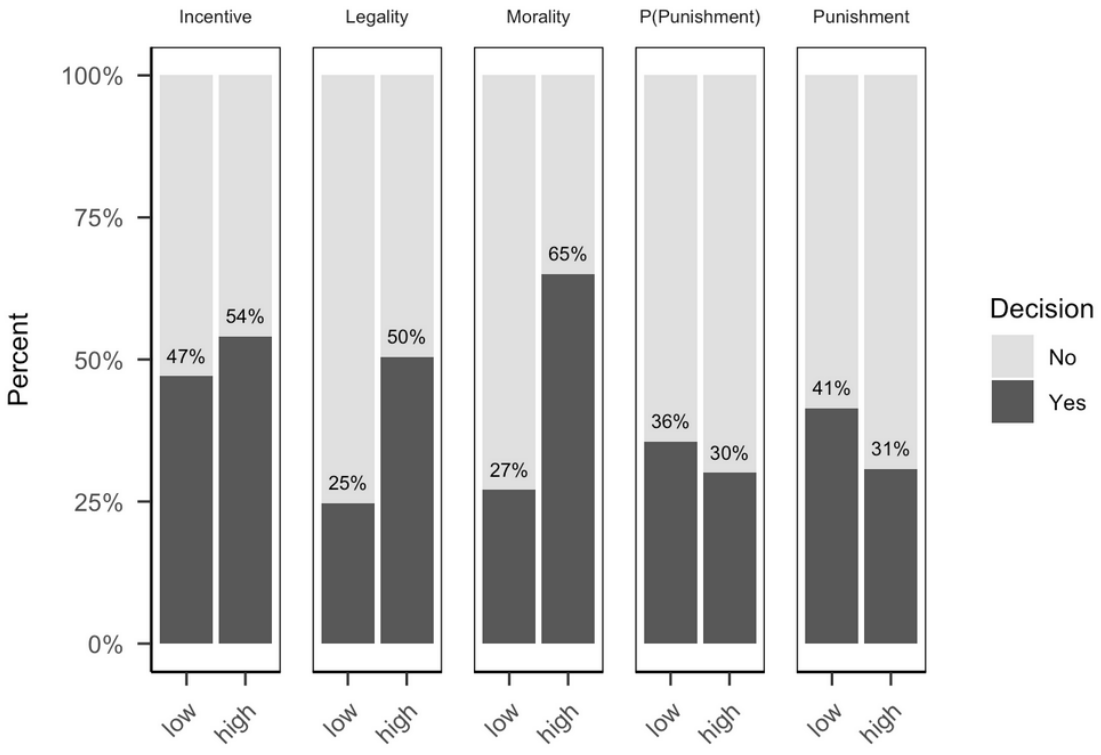
## Results

### *Dimensions' Influence on Decision Making*

As preregistered, if vignettes were read in less than 5 seconds (i.e., if participants spent less than 5 seconds on the screen displaying the vignette), the respective decisions were regarded as invalid and therefore excluded from the analysis (which was the case for 6% of the responses). As in Studies 1 and 2, participants more often decided against the opportunities for dissexual behavior (negative responses: $k = 7{,}767$, 62%) than in favor of them (positive responses: $k = 4{,}735$, 38%). Thus, the proportion of accepted opportunities for dissexual behavior in Study 3 was similar to the proportion of accepted opportunities for low-level crimes of students in Studies 1 and 2.

**Figure 7**

*Percentage of positive (i.e., dissexual) decisions by dimension*




*Note.* Low values on the dimensions *Incentive*, *Legality* and *Morality* and high values on the dimensions *Likelihood of Punishment* (P(Punishment)) and *Level of Punishment* (Punishment) were expected to lead to a lower percentage of positive responses.

For the analyses, values of the dimensions *Likelihood of Punishment* and *Level of Punishment* were coded so that high values (i.e., a low likelihood and a low level of punishment) are expected to favor dissexual behavioral decisions. Across all dimensions, in vignette versions with low values on the respective appraisal dimension, participants decided less often in favor of the dissexual behavior option than they did in versions with high values ($\chi^2(1, 12{,}502) = 310.82$, $p < .001$, $\phi = -0.16$).

The frequencies of positive (i.e., dissexual) and negative decisions by dimension are shown in Figure 7. To examine whether participants' decisions regarding dissexual behavior were influenced by the manipulated dimensions, we performed $\chi^2$-tests for each of the five dimensions. In this study, the manipulation led to significant effects on decisions in the predicted directions in all five appraisal dimensions: *Incentive* ($\chi^2(1, 1{,}870) = 8.79$, $p = .003$, $\phi = -0.07$), *Legality* ($\chi^2(1, 1{,}864) = 129.39$, $p < .001$, $\phi = -0.26$), M*orality* ($\chi^2$ $(1, 1{,}881) = 271.32$, $p <.001$, $\phi = -0.38$), *Likelihood of Punishment* ($\chi^2$ $(1, 2{,}500) = 8.30$, $p = .004$, $\phi = -0.06$) and *Level of Punishment* ($\chi^2$ $(1, 2{,}493) = 30.44$, $p < .001$, $\phi = -0.11$). Thus, Hypothesis 1 was corroborated for all five appraisal dimensions.

### *Associations between Person Factors and Decision Making*

Table 5 presents the bivariate Spearman correlations between dissexual decision making and person factors such as morality, self-control, sociosexual orientation, TSO, (sexual) arousal and demographic variables. As hypothesized, high morality and self-control scores were associated with lower tendencies to make dissexual decisions (Hypothesis 2). However, effect sizes were rather small (morality: $r = -.13$, $p = .002$; self-control: $r = -.23$, $p < .001$) and the correlation with self-control was larger than the correlation with morality ($\Delta r = -0.10$, $z = 2.34$, $p = .019$). Notably, the distribution of morality scores was extremely left skewed, with $M = 4.29$ and $SD = 0.76$ on a five-point Likert scale.

We also found small to medium correlations between dissexual decisions and sexuality-related traits such as sexual disgust (5-FES, $r = -.22$, $p < .001$), TSO ($r = .17$, $p < .001$), and sociosexual orientation (SOI-R, $r = .38$, $p < .001$), where higher values indicate a more liberal orientation towards uncommitted sex (Hypothesis 3). Of the three SOI-R subscales, *Desire* showed the strongest correlation with dissexual decisions ($r = .37$, $p < .001$). We found no or only very small significant correlations with participants' education level and age ($|rs| \leq .10$).

Sexual arousal was found to be positively correlated with dissexual decisions ($r = .40$, $p < .001$), whereas positive affect and non-sexual arousal were not or only weakly associated with decision making ($r$s $\leq .16$). Participants reported moderate to high levels of sexual arousal after completing the vignette task ($M = 3.16$, $SD = 1.19$, on a five-point Likert scale). In order to examine whether the individual positions of the vignettes (as a proxy for sexual arousal) predict dissexual decisions, a simple linear regression analysis was used. Contrary to our hypothesis, the position in the sequence of vignettes was not related to dissexual decisions ($\beta = .04$, $t = 0.80$, $p = .426$). Thus, Hypothesis 3 was only partially confirmed.

## Table 5

*Overview of correlations from Study 3 (Spearman)*

| | *M* | *SD* | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. | 13. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Dissexual Decisions [a] | 1.38 | 0.19 | -- | | | | | | | | | | | | |
| **Traits** | | | | | | | | | | | | | | | |
| 2. Morality | 4.29 | 0.76 | -.13** | -- | | | | | | | | | | | |
| 3. Self-Control | 3.07 | 0.60 | -.24** | .21** | -- | | | | | | | | | | |
| **Sexuality-related variables** | | | | | | | | | | | | | | | |
| 4. Sociosexuality (total) [b] | 3.78 | 0.70 | .38** | .01 | -.09* | -- | | | | | | | | | |
| 5. Sociosexuality- Behavior [b] | 3.23 | 1.05 | .25** | .06 | .00 | .80** | -- | | | | | | | | |
| 6. Sociosexuality - Attitude [b] | 4.35 | 0.80 | .23** | -.04 | -.08* | .69** | .42** | -- | | | | | | | |
| 7. Sociosexuality - Desire [b] | 3.77 | 0.94 | .37** | -.00 | -.13** | .69** | .27** | .30** | -- | | | | | | |
| 8. Total Sexual Outlet [c] | 9.89 | 6.05 | .17** | -.09* | -.19** | .30** | .23** | .17** | .25** | -- | | | | | |
| 9. Sexual Disgust | 1.50 | 0.56 | -.22** | -.04 | .08* | -.39** | -.30** | -.38** | -.22** | -.09* | -- | | | | |
| **Arousal** | | | | | | | | | | | | | | | |
| 10. Sexual Arousal | 3.16 | 1.19 | .40** | -.01 | -.08* | .28** | .10* | .13** | .40** | .20** | -.11 | -- | | | |
| 11. Non-sexual Arousal | 1.81 | 0.89 | .16** | -.04 | -.12** | .10** | .02 | -.04 | .22** | .07 | .07 | .44** | -- | | |
| 12. Positive Affect | 2.88 | 0.36 | .01 | -.04 | .06 | .05 | .03 | .03 | .05 | -.00 | .03 | .06 | .07 | -- | |
| **Demographics** | | | | | | | | | | | | | | | |
| 13. Education [d] | 3.96 | 1.24 | -.10** | -.07 | .12** | .01 | -.03 | .07 | .02 | .03 | .06 | -.09* | .03 | -.00 | -- |
| 14. Age | 37.33 | 10.87 | -.07 | .24** | .14** | .09* | .22** | .04 | -.11** | -.15** | -.19** | -.08* | -.05 | .01 | -.05 |

*Note*. [a] Dissexual Decisions: Mean decision across all vignettes (1=No, 2 =Yes); [b] Higher values indicate a more liberal orientation towards uncommitted sexual relationships; [c] Mean of TSO/week (number of orgasms per week within the last 6 months) and TSO/max (maximum number per week since age 15); [d] 1 = no degree, 2 = Hauptschule (basic school degree), 3 = Realschule (secondary school degree), 4 = university entrance qualification, 5 = Bachelor's degree, 6 = Master's degree, 7 = PhD; if not stated otherwise: higher values indicate higher manifestations; * $p < .05$, ** $p < .01$.

### *Interactions between Person Factors and Decision Making*

In order to examine whether participants with high values on morality and self-control were more or less sensitive towards a manipulation of the dimensions (Hypothesis 4), we conducted separate $\chi^2$-tests for participants with high and low self-control and morality scores and compared their effect sizes with each other (as we did in Studies 1 & 2 for psychopathy).

Participants whose morality scores were in the first (≤ 3.86) and fourth quartiles (=5.00) of the distribution were defined as morality low scorers ($n$ = 173) or high scorers ($n$ = 179). Note that due to the left-skewed distribution, high scorers' mean value was equal to the endpoint of the scale. The results of the $\chi^2$-tests that examined whether low and high scorers differed regarding their sensitivity towards the manipulation are displayed in Table 6.

**Table 6**

*Dimensions' influence on decision making of participants with low and high morality scores: $\chi^2$-tests.*

| Dimension | Morality | *k* | $\chi^2$ | *p* | ϕ | Δϕ |
|---|---|---|---|---|---|---|
| Across all dimensions | low | 3185 | 86.12 | .00*** | -0.17 | -0.01 |
| | high | 3234 | 77.20 | .00*** | -0.16 | |
| Incentive | low | 478 | 4.46 | .04* | -.10 | -0.09 |
| | high | 482 | 0.04 | .85 | -.01 | |
| Legality | low | 478 | 40.68 | .00*** | -.30 | -0.07 |
| | high | 480 | 25.27 | .00*** | -.23 | |
| Morality | low | 480 | 72.16 | .00*** | -.39 | 0.01 |
| | high | 488 | 75.89 | .00*** | -.40 | |
| Likelihood of Punishment | low | 635 | 5.91 | .02* | -.10 | -0.05 |
| | high | 648 | 1.51 | .22 | -.05 | |
| Level of Punishment | low | 636 | 3.57 | .06 | -.08 | 0.10 |
| | high | 641 | 19.65 | .00*** | -.18 | |

*Note.* low/high morality = scores in the first quartile (≤3.86) and in the fourth quartile (=5.00) of the distribution; *k*= number of decisions included; the three vignettes for which the manipulation check had failed were only included in the $\chi^2$-tests across all dimensions. * $p < .05$, ** $p < .01$., *** $p < .001$;

We hypothesized that participants with high morality scores should show no or only weak responses to a manipulation of the three dimensions *Incentive*, *Likelihood,* and *Level of Punishment* in comparison to participants low in morality. Although this assumption was correct for *Incentive* and *Likelihood of Punishment* (as the $\chi^2$-tests were only significant for people with low morality), the effect was opposite for *Level of Punishment* ($\chi^2$-test was only significant for people with high morality). However, across all dimensions and for every single dimension, the differences in manipulation sensitivity between morality high and low

scorers were rather small and did not reach significance ($0.01 \leq |\Delta\phi| \leq 0.10$, $p > .05$). Thus, Hypothesis 4 a) was only partially confirmed.

Results for interactions of self-control with decision making are displayed in Table 7. Again, participants were classified as low scorers if their self-control scores were in the first quartile ($\leq 2.62$, $n = 172$) and as high scorers if their self-control scores were in the fourth quartile ($\geq 3.46$, $n = 187$) of the distribution.

**Table 7**

*Dimensions' influence on decision making of participants with low and high self-control scores: $\chi^2$-tests*

| Dimension | Self-Control | *k* | $\chi^2$ | *p* | ϕ | Δϕ |
|---|---|---|---|---|---|---|
| Across all dimensions | low | 3197 | 119.18 | .00*** | -0.19 | -0.06* |
| | high | 3425 | 60.27 | .00*** | -0.13 | |
| Incentive | low | 483 | 7.78 | .01** | -.13 | -0.13* |
| | high | 515 | 0.00 | 1.00 | .00 | |
| Legality | low | 481 | 55.50 | .00*** | -.34 | -0.11 |
| | high | 509 | 26.13 | .00*** | -.23 | |
| Morality | low | 478 | 80.12 | .00*** | -.41 | 0.07 |
| | high | 518 | 56.87 | .00*** | -.34 | |
| Likelihood of Punishment | low | 634 | 4.15 | .04* | -.08 | -0.04 |
| | high | 686 | 1.02 | .31 | -.04 | |
| Level of Punishment | low | 637 | 7.98 | .01** | -.12 | 0.00 |
| | high | 678 | 9.55 | .00** | -.12 | |

*Note.* low/high self-control = scores in the first quartile ($\leq 2.62$) and the fourth quartile ($\geq 3.46$) of the distribution; *k*= number of decisions included; the three vignettes for which the manipulation check had failed were only included in the $\chi^2$-tests across all dimensions. * $p < .05$, ** $p < .01$., *** $p < .001$

Across all vignettes, participants with low self-control scores were more strongly influenced by the manipulation of the dimensions than were participants with high self-control ($\Delta\phi = -0.06$, $z = -2.50$, $p = .012$, Figure 8). On the dimensional level, the difference between low and high scorers was significant for vignettes in which *Incentive* was manipulated ($\Delta\phi = -0.13$, $z = -2.06$, $p = .040$). Participants with high self-control were not influenced at all by a manipulation of the dimension *Incentive* ($\phi = .00$); regardless of the level of *Incentive*, they had a low probability to make a dissexual decision (see Figure 9). In contrast, participants with low self-control generally showed a higher tendency to make dissexual decisions and they were influenced by the manipulation (more dissexual decisions if the *Incentive* was high than if it was low). Hypothesis 4 b) was therefore confirmed.

**Figure 8**

*Probability of a "Yes" response (i.e., dissexual decision) for participants high and low in self-control across all dimensions*

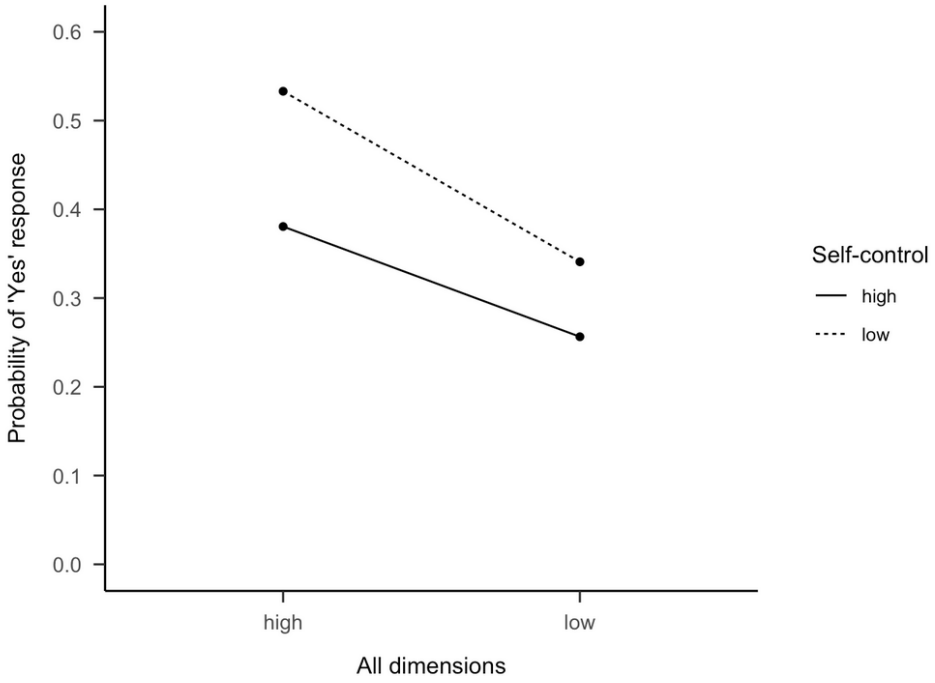
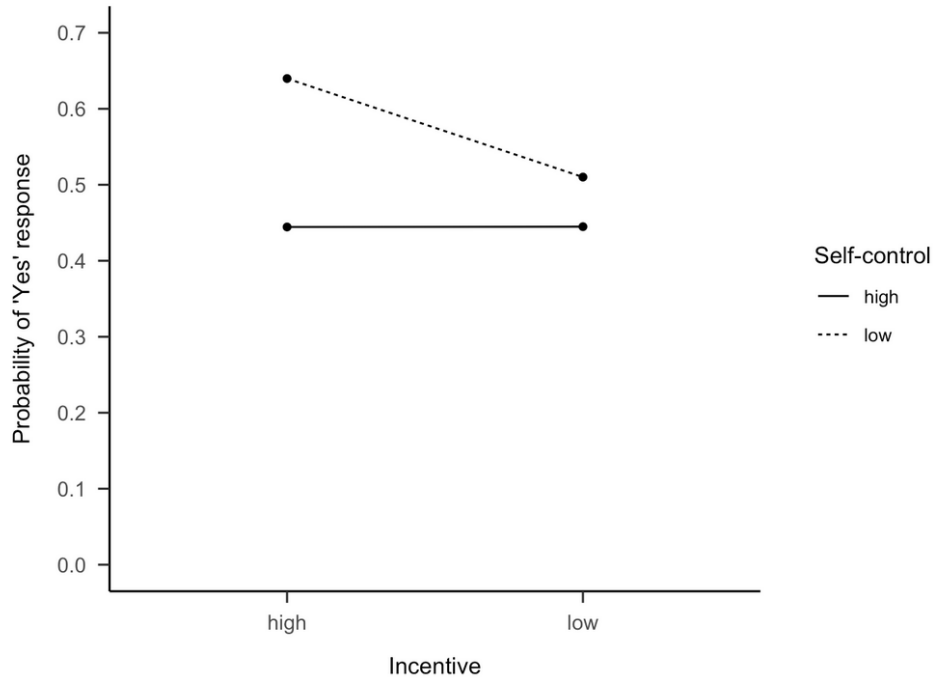



**Figure 9**

*Probability of a "Yes" response (i.e., dissexual decision) for participants high and low in self-control for vignettes in which Incentive was manipulated.*



**Discussion**

In Study 3, we replicated the findings from Studies 1 and 2 in that the dimensions *Legality*, *Morality*, *Likelihood,* and *Level of Punishment* had an impact on criminal decision making. Additionally, Study 3 showed that *Incentive* also impacted decision making in the domain of dissexual behavior. Why did we find this effect for *Incentive* in Study 3 on dissexual decision making, when we did not observe it in Studies 1 and 2 on general delinquency? One explanation could be that participants in Studies 1 and 2 were in a "cool" mode and therefore had difficulty imagining the benefits of the hypothetical actions, as is in line with the hot-cold empathy gap (Loewenstein et al., 1997). For example, in the "fake discount" vignette, participants may have had trouble imagining how badly they would have wanted a particular (very attractive vs. moderately attractive) pair of shoes in the described situation, or in the "literature download" vignette, how badly they would have needed a certain research article for their thesis (due the next day vs. in six weeks' time), and thus how strong the incentive would have been if those situations were real. In Study 3, however, most participants stated that they were at least moderately aroused after reading the vignettes, which means they were in a "hot" mode. Thus, there was no such hot-cold empathy gap and participants may have estimated the potential incentives and their potential behavior more realistically.

Dissexual behavior is a common phenomenon (Fedina et al., 2018; Finkelhor et al., 2014; Müller & Schröttle, 2004). However, because of social desirability, we did not expect participants to be completely open and honest about engaging in this kind of behaviors in a hypothetical scenario study. Surprisingly, though, we found high rates of positive decisions even for vignettes that describe behaviors that severely restrict sexual self-determination. For example, 36% said that they would continue sexual intercourse at least for a while if their sexual partner said that she wanted to stop because she no longer enjoyed it. Still 26% would continue intercourse in the other version of the vignette in which she stated that she felt pain due to a recent appendectomy (manipulation of *Morality*). At the same time, participants showed very high moral standards on the trait morality scale, which surveyed the wrongfulness of various non-sexual crimes. This illustrates that sexual arousal reduces sexual inhibition, but it does not affect the moral evaluation or execution of non-sexual antisocial behavior (Imhoff & Schmidt, 2014). Although sexual arousal was positively correlated with the frequency of dissexual decisions (as well as low sexual disgust, high sex drive and a

liberal sociosexual orientation), the position of the individual vignette in the sequence of all vignettes did not predict dissexual decisions. This result suggests that the position of the vignettes was not an appropriate proxy for sexual arousal or that participants did not become increasingly aroused during the task but were already aroused when they started the task. Since most participants were recruited via an erotic dating website, this could be a plausible conclusion.

There is also the question of why age was not associated with dissexual decisions. Generally, according to the literature, younger participants show higher delinquency rates (Farrington, 1986; Moffitt, 1993). However, this association could have been eliminated by a counteracting effect, namely that older men are less sensitive to the sexual self-determination and empowerment of women because they have internalized more traditional gender roles (Ford & Donis, 1996; Hammond et al., 2018).

In contrast to SAT, we were not able to show that morality was more crucial than self-control in decision making. In fact, the association between self-control and decision making was stronger than between morality and decision making. In our sample, the weak correlation with morality was probably caused by the limited variance in morality scores, as the distribution was skewed to the left. The hypothesis that participants with high morality scores should show no or only weak responses to a manipulation of the dimensions *Incentive*, *Likelihood,* and *Level of Punishment* compared to participants with low morality scores was only partially confirmed because in our distribution even the moral standards of "low scorers" were rather high. However, we found some evidence that *Incentive* and *Likelihood of Punishment* were relevant to low morality scorers' decisions but not to high scorer's decisions. This is in line with SAT because people with high moral standards would not even consider crime as an option due to their moral filter. When the moral filter has failed to exclude crime from the variety of action alternatives (i.e., in participants with low morality), *Incentive* and *Likelihood of Punishment* were relevant. Because the evidence in our study is rather inconclusive (as the difference between the two groups was not significant and the effect was opposite for *Level of Punishment*), this should be re-examined in future research using a sample with higher "moral flexibility".

However, we found the predicted interaction between self-control and decision making: Across all dimensions, participants with low self-control scores were more strongly

influenced by the manipulation than were participants high in self-control. This difference was mainly driven by vignettes in which the dimension *Incentive* was manipulated. Participants with high self-control showed a low base rate of endorsing (hypothetical) dissexual behavior – independent of whether the incentive was high or low. Participants with low self-control, however, had a higher probability of making dissexual decisions when the incentive was high than when it was low, showing that they were influenced by a manipulation of this dimension. According to Gottfredson and Hirschi (1990), people with low self-control are impulsive and therefore unable to resist the easy, immediate gratification that crime and analogous behaviors (such as smoking, drinking, gambling, and engaging in risky sexual relationships; van Gelder et al., 2015) almost ubiquitously offer in everyday life. Because self-control has been equated with "criminal propensity", the theory has been criticized as being tautological (Akers, 1991). Therefore it is important to understand through which processes self-control is related to crime. Nagin and Paternoster (1993) found that participants with low self-control perceived the rewards of crime as more valuable. This is consistent with our finding that participants with low self-control were more likely to make a dissexual decision if the incentive (e.g., attractiveness of the target) was high, while this was not relevant to participants with high self-control. However, in our study it remains unclear whether participants with low self-control perceive incentives as higher than other people perceive the incentives to be (i.e., differences in incentive appraisal), or if they put more weight on incentives, or even do not consider other dimensions at all.

Although Nagin and Paternoster (1993) found that people with low self-control also perceived the costs of crime as less aversive, we did not find any interactions between self-control, dissexual decision making, and *Likelihood of Punishment* or *Level of Punishment*. This ties in with a series of inconsistent findings on self-control and deterrence. Nagin and Pogarsky (2001), Piquero and Tibbetts (1996), and Pogarsky (2002) observed that effects of deterrence factors such as detection probability and expected fine were reduced in people with low self-control; Wright et al. (2004) found that they were enhanced in people with low self-control; and Waubert de Puiseau et al. (2019) found no interaction at all with self-control, as we found in our study. Therefore, the relationship between self-control, deterrence factors, and crime does not seem robust. Our data suggest that the effect is conveyed by the appraisal of the incentives rather than by the costs – at least in the domain of dissexual behavior.

## Interim Discussion

### Summary Studies 1-3

We have introduced the Appraisal Model of Criminal Decision Making which comprises six appraisal dimensions: *Incentive*, *Feasibility*, *Legality*, *Morality*, *Likelihood* and *Level of Punishment*. In order to test its basic assumptions, i.e., that the appraisal dimensions play a role in decision making and that person factors interact with these appraisals, we conducted three experimental studies in two relevant domains – low level delinquent behavior and dissexual behavior. Evidence showed that all dimensions that were tested in these studies do influence decision making. The influence of the dimension *Incentive* was only evident in Study 3, which might be explained by the hot-cold empathy gap that could have hampered the examination of this dimension in Studies 1 and 2. In both domains, the manipulation of *Morality* had the strongest impact on decisions, and was also considerably strong for participants with high psychopathy scores. We also found interactions between person factors (psychopathy in Studies 1 and 2, morality and self-control in Study 3) and certain appraisal dimensions in criminal decision making. These results provide first indications that it is possible to translate person factors into appraisal patterns, and therefore to develop a better understanding of their relationship to delinquency.

### Limitations

The question arises whether the conclusions from the present research can be deemed valid as the manipulation check showed that an orthogonal manipulation is not possible. To address this problem, we excluded the vignettes in which the two versions of the vignette did not differ significantly regarding the dimension that was intended to be manipulated according to the ratings from our independent sample. Nevertheless, many of the remaining vignettes differed significantly in more than one dimension. This is not surprising, as the dimensions are also confounded in the real world. For example, it is hardly possible to manipulate a vignette's legality without manipulating the consequences or punishment resulting from this action. However, it is striking that participants obviously were unable to evaluate the dimensions independently – even in cases where it was not a problem to manipulate them independently. For example, in a vignette that dealt with the question of whether one would cross the street at a red traffic light, we manipulated the *Likelihood of Punishment.* In version

one, the police were visibly nearby; in version two, no one was in sight. The fine itself remained unaffected by the manipulation. Nevertheless, in our manipulation check study, the level of punishment was rated as significantly different across the two versions. At first glance, this may raise doubts about the validity of the ratings, but it also sheds light on the processes involved in assessing these kinds of situations: It indicates that participants process the scenarios holistically before they rate a single dimension. Although we did not ask them whether they would carry out the action depicted in the vignette, they most likely answered this question to themselves, and were influenced by this when they rated the dimensions. To prevent this in future studies, participants should be reminded that the dimensions should be evaluated separately and independently from each other. Notably, this insight supports our approach of having the vignettes evaluated by an independent sample. In previous studies, participants have been asked to evaluate individual aspects of the situation and to make decisions (Bachman et al., 1992; Nagin & Paternoster, 1993; Paternoster & Simpson, 1996). Under these circumstances, the mutual influences of evaluations and decisions might have been very strong, as people in general strive for consistency: Participants probably struggle to rate a certain action as immoral immediately after stating that they would carry out the action and vice versa. Because an orthogonal manipulation of the dimensions is not possible and because of the resulting limitations regarding the internal validity, future research should take a regression-analytical approach instead of the experimental approach (see Study 5).

Another limitation to the research we have presented is that based on the data from Studies 1-3 we do not know whether person factors influence the appraisals themselves or only the weighting of the dimensions. For example, in the case of psychopathy, whether an action was illegal or legal was very relevant to participants who scored low in psychopathy but did not make a difference to those who scored high in psychopathy. This raises the question of whether psychopathy high scorers simply do not weight the legality aspect as strongly (i.e., they do not care if an action is illegal), or whether they have a shifted setpoint and underestimate the illegality of the action.

In addition, the extent to which the *Feasibility* dimension plays a role in decision making remains unexplored. It appeared self-evident to us that it is essential to have the necessary competencies, and that an action must be feasible in order to be carried out. Therefore, the impact of the *Feasibility* dimension was not empirically investigated here. In our reasoning, if

the feasibility is limited, the action should be carried out less frequently. However, it is conceivable that persons with certain traits may find it particularly attractive to commit offences that are difficult to execute. For the sake of completeness, the *Feasibility* dimension was examined in Studies 4 and 5.

Finally, although there is a body of evidence suggesting that the decision on hypothetical criminal behavior is a valid predictor of actual behavior, it remains a limitation of our research that we have not examined actual behavior.[10]

**Outlook Studies 4 and 5**

In studies 1-3 the focus was on testing whether the appraisal dimensions play a role in criminal and dissexual decision making. In Study 4 it was tested whether the appraisal dimensions were also mentioned in descriptions of real criminogenic situations by people who have actually committed crimes. We therefore conducted interviews with former offenders and asked them about the cognitive processes at the moment of deciding for or against committing a crime. Their responses were analyzed qualitatively.

In Study 5, we also used scenarios depicting criminogenic situations and asked participants to decide in favor of or against an action (as in Studies 1-3), However, this time we did not manipulate the dimensions, and we analyzed our data with a hierarchical regression model. The main objective of this study was to improve the methodological problems of Studies 1-3 and to investigate interactions between stable personal factors and appraisal dimensions in more detail.

## Study 4

In this study, a qualitative approach was chosen to complement the quantitative approaches of previous studies. Therefore, semi-structured interviews were conducted with former offenders in order to explore the situational appraisal processes that operated in the moment when they

[10] For a detailed discussion of the appropriateness of "intentions to offend" as an estimate of delinquency, see Fishbein and Ajzen (1975); Murray and Erickson (1987); Pogarsky (2004); Sheeran and Webb (2016).

decided in favor of or against committing a criminal act. We chose a sample of former offenders to examine whether the appraisal model is applicable to more serious crime compared to the everyday delinquent behaviors that were depicted in our vignettes. We asked the participants to describe the exact moment when they decided to commit a crime, and coded whether they mentioned the appraisal dimensions (confirmatory approach). Additionally, we examined whether other dimensions that are not featured in our model were mentioned (exploratory approach) in order to test whether the model was comprehensive. Participants were also asked to give their thoughts on a number of hypothetical situations, similar to the ones used in the vignette studies. Therefore, we investigated actual, past behavior as well as hypothetical behavior. The study was preregistered in the OSF (Appendix E).[11]

In addition to the interviews, we asked participants to complete the vignette task from Studies 1 and 2 on a computer (without collecting any personality measures). However, two participants only agreed to take part in the interviews but not in the vignette task, and several others who were willing to participate in the vignette task had difficulties concentrating (e.g., were distracted) or took a very long time to process all vignettes due to poor reading and/or language skills. With a sample of $n = 20$ participants who completed the vignette task it also made little sense to quantitatively analyze the dimensions' impact on the response behavior in the vignettes. For this reason, the vignette data were not analyzed. In future studies, the use of vignettes for forensic samples may still form a valuable instrument, but the setting should be better adapted to the sample. The formulation of the vignettes should be simpler (the current vignettes contained distractor material), a smaller number of vignettes should be used, and the total duration of the study should be shorter (no additional interview).

[11] Preregistration link: https://osf.io/9abf8

## Method

### *Participants*

The interviews were conducted in a residential home for men released from prison or on parole. A total of 22 men aged 21 to 55 ($M$ = 31.32, $SD$ = 8.56) took part in the study. The number of previous convictions ranged from 1 to 13 ($M$ = 4.85, $SD$ = 3.34).[12] The most common convictions were drug offences ($k$ = 11), violent offences ($k$ = 10), property offences ($k$ = 7), and fraud-related offences ($k$ = 6). The time spent in custody ranged from 0 to 42 months ($M$ = 12.56, $SD$ = 11.25). In cases of participants who did not indicate any prison sentences, the sentences had been suspended subject to a period of probation. Eleven participants reported having consumed alcohol at least once in the last 30 days (tobacco: $n$ = 17; cannabis: $n$ = 13; hard drugs, e.g., crystal meth, cocaine: $n$ = 4). The participants did not receive any compensation for their participation. The participation did not provide any advantages nor disadvantages for them and was completely voluntary and anonymous.

### *Materials and Procedure*

All participants were informed about the study conditions and consented to having the interview recorded on audiotape. Participants were told that the aim was to find out more about decision-making processes in the context of criminal offences. They were asked either before or after the interview to take part in the additional part of the study, in which demographics and data on criminal history and drug use were collected. Two interviewers conducted the interviews either together or alone.

Our semi-structured, problem-centered interview (Witzel & Reiter, 2012; interview guide see Appendix F) consisted of two parts: In the first part, we surveyed participants about five hypothetical scenarios [finding money (two different amounts), illegal downloading, fare dodging, buying stolen goods]. We instructed them to "think out loud", i.e., to verbalize everything that came to their mind while deciding in favor of or against the behavioral option

[12] Two participants did not provide information on their criminal history and drug use. These numbers therefore refer to a sample of $n$ = 20.

described in the scenarios. They were told that there were no right or wrong answers and that they were not supposed to justify their thoughts. If they decided against the criminal action described in the respective scenario, they were asked to explain under which conditions they would have decided in favor of the action, so that we could identify the appraisals that inhibited the participant from choosing to commit the (hypothetical) offence. If they decided that they would commit the action, we asked them about the main factor that led them to choose to do so.

In the second part of the interview, we asked the participants about an offence they had actually committed, that they still remembered well, and about which they were willing to talk to us in detail. We asked them to describe in great detail the situation and how it had come about. We explored what thoughts crossed their minds at the moment of decision and while carrying out the action, and whether any aspects were particularly relevant. This allowed us to determine which situational conditions or appraisals had been decisive in the respective situation. Participants were also interviewed about their emotions and physical reactions before and during the behavior execution. We motivated participants to report as freely as possible and tried to avoid closed or leading questions in order to prevent suggestive effects. Only in the final part of the interview, were the participants asked directly about the relevance of the appraisal dimensions in relation to the offence (e.g., “Did you think about getting caught?”, “What role did it play in the situation that the behavior was illegal?”). On average, the whole interview took $M = 30$ minutes ($SD = 6.50$, range: 19-40 minutes).

### *Data Analysis*

We conducted a qualitative content analysis following the approach of Mayring (2000). The audio recordings were independently coded by the two interviewers according to coding rules. The content categories were defined both on the basis of theory (i.e., the appraisal dimensions served as categories) and on the basis of the material (i.e., noticeable response patterns or phenomena resulting from exploratory content analysis). It was also coded whether participants’ statements were given spontaneously or at the interviewer’s request. The inter-rater reliability across all appraisal dimension categories was Cohen’s $\kappa = .80$ (Incentive: $\kappa = .90$, Feasibility: $\kappa = .76$, Legality: $\kappa = .80$, Morality: $\kappa = .84$, Likelihood of Punishment: $\kappa = .77$, Level of Punishment: $\kappa = .70$). Regarding the categorization whether the statement was

made spontaneously or not, the inter-rater reliability was κ = .87, and regarding the exploratory categories the inter-rater reliability varied between κ = .91 and κ = 1.0. Deviations between the two raters were discussed and an agreement was reached in each case.

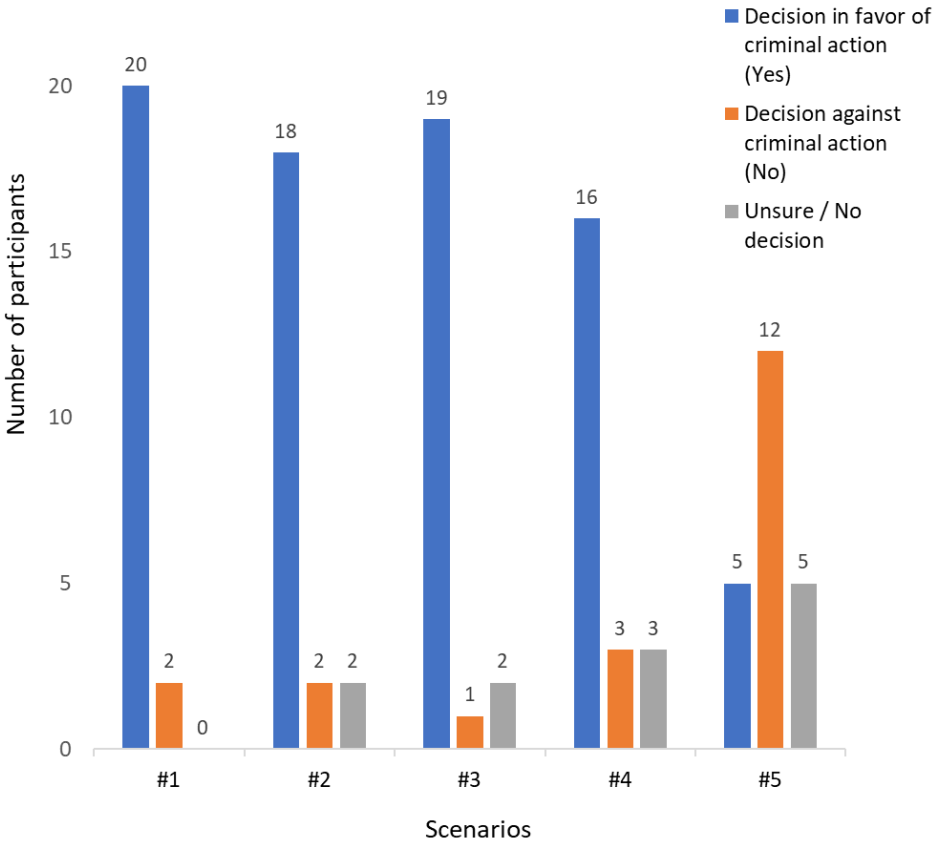
## Results and Discussion

### *Hypothetical Scenarios*

Although our focus was rather on the process of decision making than on the decisions themselves, we asked participants which decision they would make regarding the respective hypothetical actions. For Scenarios 1-4, we found a clear overall tendency to carry out the action (Figure 10). However, in Scenario 5 (buying a stolen smartphone), most participants said they would decide against it.

**Figure 10**

*Number of participants who stated that they would (not) carry out the criminal action described in the scenarios*




*Note*. Scenario #1: Keeping €20 found on the street, #2: Keeping €200 found on the street, #3: Illegal downloading, #4: Fare dodging on the tram, #5: Buying stolen smartphone. *N* = 22 for all cells.

Our primary goal was to find out whether the appraisal dimensions were considered in the process of decision making. Table 8 shows the number of participants who mentioned the respective appraisal dimension without the interviewer's request for each single scenario, and overall, i.e., the number of participants who mentioned the respective dimension in at least one scenario.

**Table 8**

*Number of participants who spontaneously named the respective dimension while being interviewed about the hypothetical scenarios*

| | Scenarios | | | | | |
|---|---|---|---|---|---|---|
| **Dimensions** | **Overall** | **#1** | **#2** | **#3** | **#4** | **#5** |
| Incentive | 22 | 9 | 12 | 12 | 13 | 13 |
| Feasibility | 3 | 0 | 0 | 3 | 0 | 0 |
| Legality | 12 | 2 | 0 | 8 | 0 | 2 |
| Morality | 20 | 19 | 11 | 7 | 2 | 3 |
| Likelihood of Punishment | 20 | 1 | 1 | 3 | 12 | 20 |
| Level of Punishment | 13 | 0 | 0 | 3 | 11 | 3 |

*Note*. Overall: Number of participants who mentioned the dimension in at least one of the scenarios. Scenario #1: Finding €20 on the street, #2: Finding €200 on the street, #3: Illegal downloading, #4: Fare dodging on the tram, #5: Buying stolen smartphone. Only mentions that were made without interviewer's request are presented. If a participant mentioned a certain dimension more than once regarding the same scenario, it was coded as one mention. $N = 22$.

Nearly all of the 22 participants mentioned the dimensions *Incentive*, *Morality* and *Likelihood of Punishment* at least once, which means that these dimensions played a role when the participants thought about executing the respective behaviors. *Legality* and *Level of Punishment* were mentioned by the majority of participants. *Feasibility* was only mentioned by three participants.

The results show that which dimensions are named strongly depends on the type of scenario. One exception was *Incentive*; this appraisal was more or less equally relevant in all scenarios. *Feasibility* was only mentioned for the "illegal downloading" scenario (#3), because some participants did not know how to download videos, whereas the actions described in the other scenarios were deemed easily feasible by all participants. Even though the *Feasibility* appraisal might have been performed for every scenario, it was only mentioned if the feasibility was limited. This indicates that appraisals are often not carried out consciously

(Ellsworth & Scherer, 2003; Scherer, 2005), and therefore might not be optimally observable in interviews. *Likelihood of Punishment* was considered especially in Scenario 5, as most participants were aware of the relatively high risk of buying a stolen smartphone because it can be identified by its IP address.

### *Offences Committed by Participants*

Since participants chose which offence they wanted to talk about, many different criminogenic situations and types of offences were addressed. Each participant was interviewed on one offence only. The most common offences selected were burglary ($k = 5$), fraud ($k = 3$), possession / trafficking of narcotics ($k = 3$), robbery ($k = 3$), assault and battery ($k = 2$). There was also one sexual offence selected. Table 9 shows how many participants mentioned the appraisal dimensions when we interviewed them about the offence from their own criminal history.

**Table 9**

*Number of participants who named the respective dimension (spontaneously and on request) while being interviewed about an offence they have committed; N = 22*

| Dimensions | Total | Spontaneously[a] | On Request[b] |
|---|---|---|---|
| Incentive | 22 | 22 | - |
| Feasibility | 19 | 8 | 11 |
| Legality | 11 | 2 | 9 |
| Morality | 18 | 10 | 8 |
| Likelihood of Punishment | 16 | 10 | 6 |
| Level of Punishment | 11 | 3 | 8 |

*Note*. [a]Spontaneously: The appraisal was described spontaneously without the interviewer asking about it. [b]On Request: The interviewer specifically asked to what extent this appraisal was relevant in the situation described. The values reflect the number of respondents who indicated that the appraisal was relevant. $N = 22$.

All 22 participants spontaneously described the incentive that was relevant to them in the respective situation. As in the hypothetical scenarios, *Feasibility* was usually not mentioned spontaneously and of the participants' own accord. However, when we asked whether it had played a role, e.g., whether the participant would have carried out the action if it had not been that easy, many participants acknowledged that this aspect did play a role. The same applied for *Legality*. *Morality* and *Likelihood of Punishment* were mentioned spontaneously by about

half of the participants, but *Level of Punishment* was only mentioned spontaneously by three participants. However, in total, all the dimensions were described by at least half of the participants.

### *Interview Quotes regarding the Dimensions*

In the following, we present a selection of interview quotes that demonstrate the relevance of the appraisal dimensions and discuss their respective context. All quotes were translated from German into English.

**Incentive.** This dimension appeared to be the most relevant because it was mentioned by each participant, both regarding the hypothetical scenarios and regarding the participants' real offences. The incentives were either financial, status-related, or hedonistic, and depended highly on the type of offence. In the following, several quotes are presented that illustrate these different kinds of incentives.

> Interviewer (I): For us, the reason to take this action would be interesting.
> Participant (P): Money. [...] If I earn 4,000 Euros per month, that's how much a lawyer earns in Germany. [...] That is definitely lucrative.
> I: Was there anything else that played a role besides the money?
> P: Nope, it was just the money.
> (P#8, 16:33)[13]

> I: Were there any other thoughts?
> P: I thought it was also pretty cool, of course. […] I used to think it was pretty gangster-like, in the nice gangster way. To buy a kilo of weed just like that. Just when you're 18, 19, 20 years old. It feels a bit cartel-like. […] That was the cool thing. I was the guy who had all the contacts.
> (P#2, 15:51, 27:00)

> I: Were there any other incentives?
> P: [...] Well, that was a group action. A collective action in which everyone had their place, in which everyone played their part. It's like when you win a soccer game together as a team, or you spend a group activity together. Except that it's a crime here. But there's a sense of belonging, clearly.
> (P#16, 15:45)

[13] Participant ID and point in the audio tape recording

The incentive did not always result from an approach motivation but rather in some instances from an avoidance motivation. For example, one participant described how he felt provoked by someone who had insulted him and commenced to physically attack the person because he felt an urge to relieve his aggression and to restore his injured pride. In this case, the reduction of negative emotions serves as an incentive. Therefore, this kind of motivation does not contradict the appraisal model's assumptions.

As the *Incentive* dimension was mentioned frequently, it can be assumed that it has a high priority and, although the order of the appraisals was of secondary relevance, *Incentive* may even function as the starting point for a sequence of appraisals. Accordingly, a behavioral option may only become relevant if the agent recognizes some kind of benefit or incentive. The importance of the *Incentive* dimension was revealed in the interview study but not in Studies 1 and 2, although partly, the same vignettes were used. In the vignette studies, we manipulated the incentives but did not find any effect on the behavioral decisions. The reason why the relevance of the incentive was emphasized in the interviews but was not found in the vignette studies could be that, in the vignette studies, participants were unable to vividly imagine the incentives. Due to the "cool", unemotional state the participants were in, they might have found it difficult to imagine how tempting the situation would have been in reality. In the interviews, we used a much smaller number of vignettes than in Studies 1 and 2 (five vs. 40 vignettes) and discussed them in greater detail, which could have made it easier for the participants to imagine the situations. Furthermore, the characteristics of the forensic sample could have played a role (e.g., forensic participants weighted incentives more strongly).

**Feasibility**. This aspect was usually not mentioned spontaneously, probably because an action will not even be considered if it is clearly not feasible. One participant illustrated the relevance of feasibility by describing that the criminal act could not be carried out as planned because the shop they intended to rob was closed.

> I: And how did you choose the shop? Were there any criteria?
> P: Actually, the plan was to rob a kiosk near my place. But on Sundays the kiosk was closed and there was a cake and pastry shop across the street, like a shop where you can buy everything. And then we just took this one.
> (P#14, 6:52)

In addition, as an exploratory category, we coded whether participants emphasized the role of the opportunity. In total, when describing the real offences they had committed, six participants highlighted the importance of the opportunity, and stated that the opportunity itself was decisive for committing the offence. A participant described that, when he entered a supermarket, he happened to see that there was an unlocked bicycle. When he came out and saw that the bicycle was still there, he took it with him.

> I: And was there a particular characteristic that the bike had that made it attractive to you? Besides the fact that the bike was not locked?
> P: No, it wasn't an overly expensive bike. Just that it was open for me was like an invitation.
> (P#6, 10:20)

Another participant who had been convicted of cybercrime experienced the need to keep buying things under a false name as a kind of addiction. When asked about the feasibility, he made the following statement:

> I: What role did feasibility or practicability play?
> P: [...] If it wasn't easily possible, it wouldn't have come so far. Or let me put it this way, if it wasn't made so easy for you. [...] In my opinion, the online traders and also the site operators should have the duty to provide more protection to the perpetrators, so that a perpetrator no longer becomes a perpetrator. [...] You open up this space to these people, although you could prevent it or, at least, make it more difficult.
> (P#3, 23:28)

This statement indicates that feasibility was an important aspect in decision making and that in the process of desistance from crime, some offenders wish there were fewer criminogenic opportunities (in his view, more thorough surveillance would not only protect the victims but also protect the perpetrators from committing further crimes). On the other hand, by making this statement, this participant protected his own self-esteem, because the blame could be attributed externally to the operators of the website. One application of the appraisal model in this participant would be for him to learn how to put less weight on feasibility and to resist impulses, even though the criminal action is very easy to perform. The relevance of opportunities has already been covered by RAT (Cohen & Felson, 1979) and is also included in our framework model that describes how stable and dynamic person factors and opportunities for criminal behavior come together (Figure 1).

**Legality.** When being interviewed about the offence they had committed, half of participants agreed that they had thought about whether or not the action was illegal, and that this aspect had mattered to them. However, contrary to our hypotheses, the fact that the action was illegal did not hinder them from committing the act but rather served as an incentive. Out of 22 participants, eight participants stated that they perceived illegality as an incentive, i.e., as something positive that made the action particularly interesting. Eleven participants said that the mere fact that something is illegal was not relevant to them apart from the fact that illegal actions might have negative consequences. Two participants felt that illegality was an incentive and an obstacle at the same time and only one participant felt it was only an obstacle. The following quotes illustrate the role of illegality as an incentive.

> I: Did you think of anything negative during the burglary?
> P: In the beginning I was thinking a little bit, hmm… this might suck [...] But the mood was right, and I thought: Fuck it. You can suppress these [negative] thoughts [...] We all know that it is forbidden. But sometimes it's also fun to do something forbidden.
> I: Was it fun because you were thinking about the money or because it was just about doing something forbidden?
> P: It was fun because I was thinking about the money, but also because it was forbidden.
> (P#15, 17:45)

> I: The fact that it was illegal, did it make it less attractive or more attractive?
> P: More attractive, I would say, because not everyone does that. It is something special to be the 1% and not like the other 99% [...] It was the incentive to do something really extreme. [...]
> I: Would there have been legal alternatives to get money?
> P: I was working, I made good money, but it was just the thrill, you know.
> (P#14, 12:48)

This surprising finding could be an example of how person factors interact with appraisals. In our vignette studies with non-forensic samples, we found that illegality had a negative effect on the probability to execute the action. However, in these previous studies we used community samples with overall high levels of self-control, low psychopathy scores, and high moral standards. It can be hypothesized that the forensic sample in the present study is different in terms of stable person factors, i.e., traits such as sensation-seeking or psychopathy, and that these traits might have influenced the appraisal and the weighting of the *Legality* dimension. Therefore, we decided to measure sensation-seeking in Study 5. A first indication

of the influence of individual difference factors on the appraisal is given by the age of the interviewees in Study 4. Participants who stated that they saw illegality as an incentive were descriptively younger than those who stated that illegality was not relevant or even hindering ($M$ = 28.20, $SD$ = 5.49, $n$ = 10 vs. $M$ = 34.25, $SD$ = 10.32, $n$ = 12; $t(20) = 1.66$, $p = .11$), which is in line with the robust finding that sensation-seeking scores decline with age (Roth et al., 2005; Zuckerman et al., 1978).

Furthermore, Studies 1 and 2 showed that psychopathy might interact with the relevance of the *Legality* dimension in decision making. We found that whether an action was illegal or legal was very relevant to people with low psychopathy scores but was not relevant to participants with higher levels of psychopathy, who showed a higher frequency of criminal decisions.

It may also be that participants found it difficult to determine what role illegality has really played. As discussed above, some dimensions are confounded with each other. Illegal acts often promise high incentives that are difficult to achieve by legal means. In addition, illegal activities usually carry the risk of being held accountable for the crime (*Likelihood of Punishment*). For people with high sensation-seeking scores, this might result in a certain kick or thrill. Up to this point, the question remains as to what extent the *Legality* dimension explains variance above and beyond the other dimensions. In the interviews, it has at least become apparent that some people have a bad gut feeling about illegal actions, even if they consider the action to be morally acceptable and of low risk, whereas others see the violation of laws as something genuinely tempting.

**Morality.** The finding that the *Morality* dimension was frequently mentioned is consistent with the results of our previous vignette studies, in which we found the largest effects for this dimension. It is also consistent with SAT (Wikström, 2010), in which crime is regarded as an act of moral rule breaking and in which a person's moral filter is the decisive mechanism for determining whether crime is an option. One participant explained that morality had always been important to him, even at times when he was committing crimes on a regular basis:

> P: Well, I have developed my own moral compass in life. I would never have stolen anything or taken anything from someone who couldn't afford it. [...] Even at the time when I was heavily addicted to heroin, I would never have ripped off grannies, for example. […]
> I: What if this [referring to a burglary] had been a family business?

> P: I knew a bit about the owner of the shop. And if this had been a family business that had no insurance [...] then I certainly wouldn't have done that. For us it was a Robin Hood story, except that we were all drug addicts.
> (P#16, 37:12)

In the scenario part of the interview, the *Morality* dimension was most often mentioned in the context of the scenarios that dealt with finding money on the street.

> I: You just said that if you saw someone looking for the €200 note, you would talk to them. Why?
> P: Yes, definitely. Because I think it is the right thing to do. If something would fall out of my pocket, and someone would see it, then I would be glad if he would tell me. And I would do that, too. [...] I'm also currently working in interior cleaning and recently I found a relatively expensive ring on the floor of a car. I almost vacuumed it away, but then I saved it and gave it to the driver, although I could have kept it. It would not have been noticed. But I did not. [...] When I see that the €200 note can be traced back to someone else, then I would also feel like shit if I did that, and I would have an extremely bad conscience.
> (P#2, 1:47)

Another participant said that he would take the money and give it to someone in need, e.g., a homeless person ("If I find a €20 note that does not belong to me, that I have not worked for or anything, then I give it to someone who needs it more than I do", P#12, 00:20). As socially desirable responding often occurs in face-to-face interview studies when sensitive personal information is addressed (Richman et al., 1999), socially desirable responding might also have been an issue in our interviews, especially when the *Morality* dimension was discussed. However, we gained the impression that the majority of participants were highly sincere in their answers. An indication of the high degree of openness of the participants is the fact that participants frequently reported offences that have not yet come to the attention of authorities and for which they could still be held accountable. Additionally, rates of criminal decisions in the hypothetical scenarios were also relatively high, although most participants, in case of rule violations, would have to fear parole revocation and/or the loss of their place in the residential home. Furthermore, emphasizing one's own moral integrity in the interviews might not have had the function of meeting social norms, but rather of maintaining a positive self-image. Whether the participants would in reality act according to the high moral standards they stated in the interviews is unclear. However, as higher scores on measures of social desirability and impression management are related to lower recidivism (Mills et al., 2003), people who respond in a socially desirable manner may also show a stronger adherence to social norms in

real life. Thus, the frequent mention of the *Morality* dimension is an indication that most participants do think about morality when they decide in favor of or against a certain action.

**Likelihood of Punishment.** The frequency of spontaneous mentions of this dimension varied relatively strongly between the scenarios, presumably because the objective probability of being caught also varied considerably. Overall, the dimension was mentioned frequently, both in the hypothetical scenarios ($n$ = 20) and while discussing the participants' real offences ($n$ = 16).

> I: Weren't you afraid you'd get caught fare dodging [on the train]?
> P: If you take this route every day, you know it well. Of course, the moment of surprise can come. But usually you know the times when someone checks. No conductor checks at 7 a.m. That's rush hour, everyone goes to work. They start checking at 6 pm, that's been my experience. I always buy a ticket after 6 pm.
> (P#8, 9:10)

The quote shows that the participant avoided the prohibited action (i.e., being on the train without a ticket) if the likelihood of punishment was high (i.e., after 6 pm). For this participant, the dimension *Likelihood of Punishment* was crucial for the decision. Another participant described the fear of being discovered metaphorically:

> I: Did you think about getting caught?
> P: Yes and no. [...] No matter how euphoric you feel in that moment, you always have a warning sign somewhere in the back of your head, flashing and saying: something could happen. But in this situation, it was only flashing a little bit.
> (P#15, 26:20)

In this quote, Participant 15 was describing a spontaneous burglary into an apartment where one of the windows was tilted open. He took the likelihood of negative consequences into account when he decided to break into the apartment. The *Likelihood of Punishment* appraisal was made, and it was acknowledged that there was a certain risk of detection. However, the action was still executed. He probably underestimated the risk (warning sign was "only flashing a little bit") because of his shifted individual setpoint in this situation. This participant emphasized several times throughout the interview that one crucial factor to break into the apartment had been that his "mood had been right". The excitement and euphoria might have led to a different appraisal and, finally, to the disregard of the warning in his head.

As his description was rather metaphorical, it could have been more of a bad gut feeling than a clear thought, which indicates that appraisals are often unconscious.

Furthermore, we observed that many participants drew a strong distinction between their thoughts before and after making the decision to carry out the action. They explained that while they were still weighing on whether or not they wanted to commit the crime, they were considering the likelihood of being caught and suffering negative consequences. However, once the decision to carry out the act had been made, these thoughts about negative consequences were deliberately suppressed. Because this phenomenon was described several times, we introduced it as a coding category. In total, nine participants described that they tried not to think of negative consequences while carrying out the act, although they had considered them before.

> I: Did you think that you could go to prison as a consequence?
> P: I didn't think about it at that moment. Nobody thinks about it at that moment. […] I try to tell myself that everything is alright. I believe that anyone who does something criminal only tries to see the positive in that situation. "It's gonna be okay. I can sell the stuff, etc." A bank robber doesn't go into a bank with the thought of getting caught. [...] At that moment I don't think about the drawbacks, "I'll go to jail, I'll be in custody, and so on". In that moment I think, "Okay, I'll have the stuff in a minute, I'll be out with the guys, we'll go for a drink, we'll laugh in a minute." [...] Because if I think of the negative, I get paranoid. And if you are paranoid, that's when you make the biggest mistakes.
> (P#15, 27:50)

Thoughts of potential punishment were seen as counterproductive for the successful execution of the act by several participants. One of them stated: “I had thought a little about the consequences. But then I thought to myself: If I am thinking about getting caught all the time, I will get caught.” (P#14, 14:42). Another one explained: “When stealing, it’s more important not to draw attention to yourself. Of course, if you’re scared, people will notice.” (P#6, 13:50). As a consequence, he tried to suppress these negative thoughts. The concern that things might go wrong if one thinks too much about the consequences was sometimes formulated as a general wisdom: “I never go into these things with a bad feeling. You should never do that.” (P#21, 17:05; see also Pb#15, 27:50, quote above). Furthermore, Participant 18 stated that thinking about potential consequences while committing a crime would be a “killjoy” and should be avoided (30:20).

**Level of Punishment.** About half of participants stated that this dimension had influenced their decision to commit a previous crime or would influence their decision in a hypothetical scenario (Tables 8 and 9). However, apart from the *Legality* dimension, *Level of Punishment* was one of the dimensions that was mentioned least. This finding is consistent with the meta-analytic finding that the relationship between punishment severity and crime is rather weak (Pratt et al., 2006) and with the finding that, historically, harsh punishment policies have not led to a decrease in crime rates (Andrews & Bonta, 2010).

When asked whether he had thought about a potential prison sentence, Participant 17 stated: "No. I probably knew I could go to jail for that. But I just didn't think about it at the time. And I'm also not sure whether I wouldn't have done it if I had thought about it." (P#17, 18:53). Another participant explained that he knew that he could face a prison sentence but was not deterred by it: "Prison is bad, but life goes on in prison. It's not nice, and nobody wants to go to jail voluntarily, but that doesn't really scare you off. I've never been scared off by that." (P#6, 12:05). Other participants stated that they did think about the punishment and considered it when they made their choice, however, they underestimated the severity of the punishment: "If I had really known that I would go to jail for four months, without parole, without everything, […] I would definitely not have done it." (P#6, 22:14.). „A buddy who also sprays graffiti has already caused €100,000 of damage in the city, and he has never received anything worse than a fine. No prison sentence or anything like that. I was sure that I would only get a fine." (P#4, 12:23). Despite formal punishment (e.g., monetary fines, prison sentences), participants stated that informal consequences played an important role (e.g., shame, disappointment of family members, etc.).

During the interviews, we noticed that some participants weighed the costs and benefits of their actions rationally. Therefore, we added the coding category "cost-benefit analysis". Cost-benefit analyses were articulated by twelve participants. For example, Participant 16 stated that, before he committed a burglary, he had thought a great deal about the level of a potential punishment and weighed it against the benefits of the action:

> P: In court it would have been treated as theft by housebreaking, in the worst case as grand theft by housebreaking. That would mean you would be in prison for a maximum of five years. But I was never caught before, and no one would give you five years the first time [...]. That means you would have been in for two to two-and-a-half years. [...] If you get released early, this means 18 months maximum. And it paid

> off well, so I was okay with that. This was my calculation.
> (P#16, 28:20)

In contrast, several participants stated that, because of their negative experiences in prison, further imprisonment would not be worthwhile in any case. Participant 14 explained how much he had missed his family and that this aspect was so important to him that it prevented him from committing further offences:

> I've celebrated Christmas, New Year's Eve, two of them, alone, and that sucks. My whole family is at home, eating, drinking, laughing. And I'm rotting in my cell, eight square meters. The loneliness. You're there with other people too, but they're all criminals, you know. And family is just indispensable. (P#14, 20:27)

**Conclusion**

In the previous vignette studies, we only asked participants for the decision regarding a potential behavior execution, but we did not learn anything about the appraisals themselves. In the current study, we presented largely similar scenarios, but gained insight into participants' thoughts. For this purpose, we interviewed a sample that is actually relevant in the law enforcement context, since participants had already committed crimes and were potentially at risk of recidivism. Besides current appraisals of hypothetical crimes, we also took a retrospective approach and interviewed participants about the appraisals made at the time of their respective past offences.

The interviews revealed that all of the appraisal dimensions play a role in decision making, but in general, some dimensions (e.g., *Incentive*) were mentioned more frequently than others (e.g., *Level of Punishment*). Which dimensions were relevant depended considerably on the context and varied across our hypothetical scenarios and across the different types of offences participants had committed. We did not identify any additional appraisal dimension that could be added to our model.

Our results suggest that stable or dynamic person factors interact with appraisal dimensions. For example, some participants appeared to have shifted appraisal setpoints in a sense that they underestimated the likelihood and/or level of consequences. Other participants made clear that even small incentives were very important to them (i.e., they were weighted strongly). Some participants described offences that were committed in a state of acute intoxication. In these cases, only few appraisals were described. One participant had robbed a

gas station after having consumed amphetamines for three days in a row, feeling like a "zombie". In this situation, the only incentive was to gain money, and no other considerations were made. Although one could argue that the appraisal model is still valid because most appraisal dimensions were not weighted at all, it might be somewhat overcomplex and not well suited to explain these types of actions.

Interestingly, we found that the *Legality* dimension was not considered at all by many participants, and in case it was considered, it was often entered into the equation with a negative sign. That means, the fact that an action was illegal favored behavior execution. The way legality is considered in decision making may be determined by participants' personality traits. First indications of an interaction between psychopathy and the *Legality* dimension was found in Studies 1 and 2. In Study 5 potential interactions between personality and appraisals will be systematically examined.

## Study 5

Study 5 resembled Studies 1 and 2 in that it also featured vignettes dealing with low-level crimes in a community sample. In contrast, however, we did not manipulate any dimensions because previous studies showed that orthogonal manipulations were hardly possible. Instead, a correlative, regression-analytical design was chosen. The current focus was on measuring interactions between stable person factors and appraisal dimensions.

In the pre-study we asked an independent sample of participants to rate all vignettes. However, this time, participants were carefully instructed about the dimensions and the rating task. In the main study, participants from another sample were asked about their behavioral decision regarding these vignettes, and additionally, personality variables (i.e., self-control, sensation-seeking and psychopathy) were assessed. The normative values generated in the pre-study were analyzed together with the decision values in a hierarchical regression model. This way it was examined to what extent the appraisal dimensions individually and in interaction with personality factors influence decision making.

In the present study, we tested the simple additive model that was presented in the introduction of this thesis. However, the assumption that a certain threshold value must be exceeded in order for the behavior to be executed was not examined in the present study because we measured the behavioral decisions in terms of probabilities instead of using a

binary decision format. Thus, we only tested the assumption that the higher the values of the weighted appraisals, the higher the probability that the behavior will be executed. Furthermore, in the present study we assumed that the appraisals are processed in parallel, although other processing modes are conceivable.

**Methodological Problems in Previous Studies**

There have been prior studies that have attempted to investigate how personality factors and situational evaluations interact in criminal decision making. In the introduction of this thesis, some of these studies are described.

Nagin and Paternoster (1993), for example, investigated how self-control interacted with the evaluation of proximate situational influences in criminal decision making. They assessed participants' self-control, presented them with three vignettes that described criminogenic situations (theft, drunk driving, sexual assault) and asked them about the probability that they would commit the action described in the vignettes. In addition, participants were asked to evaluate certain situational factors (e.g., attractiveness of the crime target, perceptions of costs and benefits of the crime). The results showed that self-control as well as the evaluation of situational characteristics were directly related to participants' intentions to commit the offence. However, self-control was also indirectly related to the probability of committing the crime by influencing participants' evaluations of situational factors. Participants with low self-control showed stronger intentions to offend, because they perceived the rewards of crime as more valuable and the costs of crime as less aversive. Similar scenario studies have followed (Bachman et al., 1992; Nagin & Paternoster, 1994; Paternoster & Simpson, 1996; van Gelder & Vries, 2012, 2014).

These prior studies show that it makes sense to combine situational and personality-based approaches in the explanation of criminal behavior. Their results suggest that cognitive appraisals may build a link between person factors and behavioral decisions.

However, there are some methodological problems that limit the validity of these studies. All studies mentioned above have used at most a handful of scenarios, and in most cases, they have analyzed each scenario separately. Therefore, the results were limited to the specific characteristics of single situations. Additionally, in all of these studies, participants were asked to estimate both their probability of carrying out the action as well as their evaluation of

situational factors (e.g., costs, benefits). These estimations might have biased each other because participants usually try to provide consistent responses.

In Studies 1-3 we addressed some of these problems. We used a larger number of vignettes (20 manipulated vignettes in total) and refrained from having the vignettes evaluated by the same sample that made the decisions. We manipulated specific situational characteristics (i.e., the appraisal dimensions) in the vignettes and tested whether these manipulations affected behavioral decisions. The results showed that the appraisal dimensions play an important role in criminal decision making, and they also indicated that there might be interactions between personality factors and appraisal dimensions.

However, a manipulation check on the vignette pairs revealed that an orthogonal manipulation of the dimensions was hardly possible, as the dimensions are also confounded in real life. This major limitation calls into question the validity of our findings. Therefore, we chose a regression-analytical approach for the present study to more precisely estimate the direct influence of the appraisal dimensions and their interactions with personality factors on decision making.

**The Present Study**

We created a large number of vignettes to generate a high variability in the relevant appraisal spaces. The vignettes depicted opportunities for different types of low-level criminal behavior. Participants were asked to indicate the probability that they would show the critical behavior. Additionally, we assessed different criminogenic personality traits (self-control, psychopathy, sensation-seeking) using self-report questionnaires, and demographics such as age and gender. In order to avoid circular correlations, i.e., that behavioral decisions are influenced by asking participants about the appraisal and vice versa, we conducted a pre-study in which all vignettes were rated with regard to the six appraisal dimensions by an independent sample. We determined an appraisal profile for each vignette based on the pre-study rating scores and used them as normative predictor values in our regression analysis.

According to the appraisal model, behavioral decisions are influenced by the six appraisal dimensions. So, our first hypothesis was that participants' decisions could be predicted by the appraisal profile of a vignette. However, due to individual differences, participants were expected to vary in the extent to which their decisions were influenced by each appraisal

dimension. In the appraisal model, this is reflected in the appraisal dimensions' weights, i.e., the personal relevance of the appraisal dimensions that is dependent on stable dispositions (i.e., personality traits), as well as the current state of the person. Thus, second, we tested whether person factors may account for differences in the weighting of the appraisal dimensions. Therefore, we examined three interaction effects:

Hypothesis 1: We expected individuals low in self-control to give more weight to the *Incentive* dimension than people high in self-control did. This expectation is consistent with the finding from Study 3 that participants with low self-control were more likely to make a dissexual decision if the incentive (e.g., attractiveness of the target) was high, while this was not relevant to participants with high self-control, and is also consistent with findings by other authors (Gottfredson & Hirschi, 1990; Nagin & Paternoster, 1993).

Hypothesis 2: We expected individuals high in sensation-seeking to give less weight to the *Likelihood of Punishment* than individuals low in sensation-seeking did. Previous evidence has shown that that high-sensation-seekers appraise their environment as less threatening (Roberti, 2004; Zuckerman et al., 1978; Zuckerman, 1994).

Hypothesis 3: Individuals scoring high in psychopathy were expected to give less weight to the *Morality* dimension than individuals scoring low in psychopathy did. This is in line with evidence that individuals high in psychopathy generally show lower levels of moral reasoning and are more oriented to egoistic concerns (Trevethan & Walker, 1989; see also Blair, 1995).

There might be more interaction effects between personality factors and appraisal processes that are not further specified. Therefore, we also tested the direct influence of personality factors on participants' general decision tendency.[14] In order to simultaneously analyze all vignettes (which previous studies did not), we used a hierarchical regression model to test our

[14] We mistakenly preregistered that we planned to only test the effects of age and gender on the general decision tendency. However, also the main effects of self-control, psychopathy, and sensation-seeking were examined in the regression model because we also tested their interaction effects. These tests were already part of the registered model, however, we missed to add their formal hypotheses.

hypotheses (with 80 vignettes forming one level that was nested in the person level of the analysis). The study was preregistered in the Open Science Framework (Appendix G).[15]

## Method

### *Participants*

Based on an a priori power analysis of simulated data (90% power[16]), we aimed for a sample size of $N = 200$ participants, 50% male, aged between 18 and 35 years (based on the age-crime curve; Moffitt, 1993). Participants were recruited via social media and personal contacts. They were offered the opportunity to participate in a lottery to win one of three €25 Amazon vouchers. Recruitment was stopped when the critical sample size was reached. From the total sample of $N = 200$ participants, two male participants were excluded because they stated that they had not taken part seriously or they showed indications of careless responding (i.e., overlooked reverse coded items). The final sample consisted of $N = 198$ participants (98 men, 100 women; age: $M = 26.65$, $SD = 5.37$). Twelve participants indicated being either younger than 18 or older than 35. As preregistered, we calculated and reported analyses with and without these participants. From the final sample, 3% stated they had either no degree or only a basic school degree (Hauptschule), 6% a secondary school degree (Realschule), 37% a university entrance qualification (Abitur), and 55% a university degree. The majority of participants were in relationship (47% unmarried, 12% married), and 41% were single. Regarding drug consumption, 81% reported having consumed alcohol at least once in the last 30 days (tobacco: 28%, cannabis: 14%, hard drugs: 3%). Five participants stated they had a criminal record.

---

[15] Preregistration link: https://osf.io/n3s6d/.

[16] For the data simulation we used standardized regression coefficients of $\beta = 0.2$ for Level 1 effects and $\beta = 0.15$ for Level 2 effects. We repeated the simulation 1000 times with varying values for $N$ and determined the number of observations which were needed to detect all of the specified effects with a probability of at least 90%.

### *Materials and Design*

**Vignettes.** We presented participants with 80 vignettes depicting everyday criminogenic situations, such as fare dodging on a train, illegal downloading, minor thefts, and insurance fraud. For example, "After a long night out, you're walking home. You still have a long way to go and you don't have money for a taxi. All of a sudden you see an old bicycle on the side of the road that is not locked. There are no other people in sight. Do you take that bike?" About 75% of the vignettes described illegal behavior; the remaining vignettes featured behavior that was not illegal, but mostly antisocial (e.g., cheating at a game). Some behaviors fell into a legal gray area in Germany. Participants were asked to estimate the probability of whether they would execute the behavior on a visual analogue scale (slider) ranging from 0% = *Definitely not* to 100% = *Definitely yes*. We opted for a continuous instead of a binary ("yes" vs. "no") measure because interaction effects are easier to interpret in linear regression models (Ai & Norton, 2003), and in order to capture uncertainties in participants' responses.

***Pre-study.*** In order to obtain normative values for each vignette regarding the six appraisal dimensions (i.e., "appraisal profiles"), we asked an independent sample of $N = 40$ participants from the Prolific Academic participant pool to rate 84 vignettes (for preregistration, see Appendix G).[17] To ensure that this sample was comparable to the main study's sample, we used the following of Prolific Academic's prescreening filters: age: 18-35 (which resulted in $M = 26.83$, $SD = 4.63$); balanced gender-ratio (50% men, 50% women); and nationality, current country of residence, first language: German. At $N = 40$ participants, our preregistered stopping rule for data collection was met, i.e., Cronbach's alpha exceeded $\alpha = .80$ for each dimension (results were $.90 \leq \alpha \leq .98$). To prevent fatigue, each participant only rated 50% of the vignettes. They were presented in a randomized order. Therefore, each vignette was rated by $19 \leq n \leq 21$ participants. All dimensions except *Legality* were rated on a visual analogue scale (slider) from 0 to 100 (e.g., 0 = *morally not questionable at all*, 100 = *morally very questionable*). *Legality* was measured as a binary categorial variable (*legally permitted* vs. *forbidden*). In our previous studies, participants had found it difficult to rate each dimension

[17] Pre-study preregistration link: https://osf.io/kh9yj.

independent of the other dimensions (e.g., estimating *Level of Punishment* without considering its likelihood), so they were now carefully instructed to do so, and received detailed definitions of the dimensions in combination with a training vignette. The same instructions were shown as a reminder after participants had completed 50% of the vignette ratings. Each participant received €4,40 as compensation.

Pre-study results, i.e., the appraisal profiles for each vignette, can be found in Appendix H. As we aimed for a total of 80 vignettes, we excluded four vignettes for the main study. One vignette was excluded because of a programming error, alongside three other vignettes for which the ratings had differed the most (highest sum of *SDs* across all dimensions).

**Other measures.**

***Self-control.*** To assess participants' self-control capacities we used the German short version of the Self-Control Scale by Bertrams and Dickhäuser (2009, SCS-K-D; English version by Tangney et al., 2004). The scale consisted of 13 items (e.g., "I am lazy," $\alpha = .83$) and responses were made on a five-point Likert scale (1 = *not at all*; 5 = *very much*).

***Dark Triad.*** We used the German version of the Short Dark Triad questionnaire (SD3; Malesza et al., 2019; English version by Jones & Paulhus, 2014) in order to measure psychopathy (e.g., "People who mess with me always regret it.", $\alpha = .71$), narcissism (e.g., "I insist on getting the respect I deserve," $\alpha = .68$), and Machiavellianism (e.g., "I like to use clever manipulation to get my way," $\alpha = .79$), with nine items each. Participants indicated their level of agreement on a five-point Likert scale (1 = *strongly disagree*, 5 = *strongly agree*). The SD3 was chosen because it offers an economical way to measure psychopathy[18]. Machiavellianism and narcissism were assessed for exploratory reasons because they are predictors of criminal behavior and misconduct (Azizli et al., 2016; Blickle et al., 2006; Hepper et al., 2014).

[18] We are aware that the SD3 as a short economic instrument cannot fully capture psychopathy as a complex personality construct. However, a validation study by Jones and Paulhus (2014) revealed a correlation of $r = .78$ (disattenuated correlation: $r = .92$) between the SD3 psychopathy scale and the more established Self-Report Psychopathy scale, third version (SRP-III; 64 items; Paulhus et al., 2015).

***Sensation-Seeking.*** To assess sensation-seeking, we used the *Need for Stimulation* subscale from the Need Inventory of Sensation Seeking (NISS; Roth & Hammelstein, 2012; German version), which consists of 11 items (e.g., "I like the feeling of excitement in my body," α = .91). Using a Likert scale (1 = *almost never*; 2 = *seldom*; 3 = *occasionally*; 4 = *frequently;* 5 = *almost always*), participants indicated how often in the past six months they had felt as described in each item.

***Demographics and Delinquent Lifestyle Information.*** Additionally, we asked participants to indicate their gender, age, education level, marital status, previous convictions, and drug consumption during the last 30 days (alcohol; tobacco; cannabis; hard drugs, e.g., crystal meth, cocaine).

### *Procedure*

At the beginning, participants were informed that the study was on "decision making in different situations" and were asked for their consent. They were then instructed about the vignette task, i.e., they were informed that there were no right or wrong answers and asked to indicate the probability that they would carry out the action described in the scenario. The vignettes were presented in a randomized order. After the vignettes, the other measures were displayed in the following order: SCS-K-D, SD3, NISS, demographics, information on drug consumption and previous convictions, and the seriousness check item by Aust et al. (2013). Finally, participants were asked if they wanted to make any further comments on the study.

### *Data Analysis and Regression Model Description*

In order to test our hypotheses and predict participants' behavioral decisions, we used a hierarchical regression model. The rating scores from the pre-study were averaged for each appraisal dimension and combined with the main study's data set. Because of the repeated measures design, there were two levels of variation in the regression model: vignettes (*v*, Level 1, within participants variation), nested within participants (*p*, Level 2, between participants variation).

In the Level 1 model, we tested whether the appraisal ratings of the vignettes (from the pre-study) predicted the participants' decision making (probability ratings from the main study). The Level 1 model can be formalized as follows:

$$Decision_{vp} = \beta_{0p} + \beta_{1p} \cdot Incentive_v + \beta_{2p} \cdot Feasibility_v + \beta_{3p} \cdot Legality_v + \beta_{4p} \cdot Morality_v + \beta_{5p} \cdot P(Punishment)_v + \beta_{6p} \cdot Punishment_v + \varepsilon_{vp}$$

The intercept $\beta_{0p}$ represents the general decision tendency of a participant and the slopes $\beta_{1-6p}$ represent the weights of the appraisal dimensions in the decision.[19] $\varepsilon_{vp}$ represents the error term. For *Incentive*, *Feasibility*, and *Legality*, we expected slopes greater than zero, because higher values on these dimensions were expected to increase the probability of carrying out the behavior. For *Morality* (which was reverse coded), *Likelihood of Punishment* [i.e., P(Punishment)], and *Level of Punishment* (i.e., Punishment), we expected slopes below zero.

Based on our theoretical model, we expected participants to vary regarding the weighting of the appraisal dimensions, i.e., we expected the slopes of the Level 1 model to vary between participants. We used Level 2 information (i.e., person factors) to explain these differences between participants. For each slope, we therefore specified a Level 2 model (person-level, fixed term: $\gamma_{i0}$) and added a random term ($\upsilon_{ip}$) to it. As described in our hypotheses, we expected psychopathy to account for individual differences in the slope of *Morality*, self-control for differences in the slope of *Incentive,* and sensation-seeking for differences in the slope of *Likelihood of Punishment*. Therefore, these three personality variables were added as predictors to the Level 2 equations. All other slopes were treated as random effects without any predictor variables:

$$\beta_{1p} = \gamma_{10} + \gamma_{11} \cdot \text{self-control} + \upsilon_{1p}$$

$$\beta_{2p} = \gamma_{20} + \upsilon_{2p}$$

$$\beta_{3p} = \gamma_{30} + \upsilon_{3p}$$

$$\beta_{4p} = \gamma_{40} + \gamma_{41} \cdot \text{psychopathy} + \upsilon_{4p}$$

$$\beta_{5p} = \gamma_{50} + \gamma_{51} \cdot \text{sensation-seeking} + \upsilon_{5p}$$

$$\beta_{6p} = \gamma_{60} + \upsilon_{6p}$$

---

[19] In the model equations, "β" represents the true effects. However, in the results section, "β" represents standardized regression coefficients.

Because person factors may also impact (criminal) decisions independent of the appraisal dimensions, i.e., either directly or through processes that have not been included in our appraisal model, we allowed the intercept of the Level 1 model ($\beta_{0p}$) to vary between participants. To predict this intercept variation, we used the demographic variables gender and age, and also the personality variables self-control, psychopathy, and sensation-seeking:[20]

$$\beta_{0p} = \gamma_{00} + \gamma_{01} \cdot \text{gender} + \gamma_{02} \cdot \text{age} + \gamma_{03} \cdot \text{self-control} + \gamma_{04} \cdot \text{psychopathy} + \gamma_{05} \cdot \text{sensation-seeking} + \upsilon_{0p}$$

After integrating these equations, the complete model was formalized as follows:[21]

$$\text{Decision}_{vp} = \gamma_{00} + \gamma_{01} \cdot \text{gender} + \gamma_{02} \cdot \text{age} + \gamma_{03} \cdot \text{self-control} + \gamma_{04} \cdot \text{psychopathy} + \gamma_{05} \cdot \text{sensation-seeking} + \gamma_{10} \cdot \text{Incentive} + \gamma 11 \cdot \text{self-control} \cdot \text{Incentive} + \gamma_{20} \cdot \text{Feasibility} + \gamma_{30} \cdot \text{Legality} + \gamma_{40} \cdot \text{Morality} + \gamma_{41} \cdot \text{psychopathy} \cdot \text{Morality} + \gamma_{50} \cdot \text{P(Punishment)} + \gamma_{51} \cdot \text{P(Punishment)} \cdot \text{sensation-seeking} + \gamma_{60} \cdot \text{Punishment} + \upsilon_{0p} + \upsilon_{1p} \cdot \text{Incentive} + \upsilon_{2p} \cdot \text{Feasibility} + \upsilon_{3p} \cdot \text{Legality} + \upsilon_{4p} \cdot \text{Morality} + \upsilon_{5p} \cdot \text{P(Punishment)} + \upsilon_{6p} \cdot \text{Punishment} + \varepsilon_{vp}$$

Before the model was fitted, all variables were centered (group-mean-centering for Level 1 predictors, grand-mean centering for Level 2 predictors; see recommendations for testing cross-level interactions by Aguinis et al., 2013). The R-Code for the regression analysis was preregistered based on simulated data and is accessible at https://osf.io/n3s6d/.

## Results

### *Descriptive Statistics*

Bivariate Spearman correlations between the appraisals regarding the six dimensions (from the pre-study) and participants' decisions (from the main study) are displayed in Table 10. All appraisal dimensions correlated with the decision to carry out the (criminal) action in the expected directions ($|.30| \leq r \leq |.83|$). The strongest correlation was found between the

[20] $\gamma_{00}$ reflects the overall mean in decision making across all vignettes and participants and $\upsilon_{0p}$ describes the random non-explainable difference between the overall mean and the participants' individual values.

[21] In the preregistration, the fixed effect terms for self-control, psychopathy and sensation-seeking were accidently omitted.

decision outcome and *Morality* (i.e., high moral concerns lead to an inhibition to carry out the act), and weakest were found between the decision outcome and *Feasibility*. Overall, participants showed a tendency to reject rather than accept the criminogenic opportunities depicted in the vignettes (Overall probability of carrying out the actions: $M$ = 38%, $SD$ = 21%).

In some cases, there were high intercorrelations between the appraisal dimensions (e.g., *Level of Punishment* was positively correlated with *Likelihood of Punishment*, $r$ = .80; and negatively correlated with Legality, $r$ = -.73).

**Table 10**

*Means (M), standard deviations (SD), and Spearman Correlations of the appraisal dimensions and positive decisions, i.e., in favor of the (criminal) action*

| | *M* | *SD* | 1. | 2. | 3. | 4. | 5. | 6. |
|---|---|---|---|---|---|---|---|---|
| 1. Decisions | 0.38 | 0.21 | -- | | | | | |
| 2. Incentive | 71.85 | 17.59 | .34** | -- | | | | |
| 3. Feasibility | 84.60 | 13.52 | .30** | .06 | -- | | | |
| 4. Legality | 32.54 | 38.98 | .34** | -.14 | .05 | -- | | |
| 5. Morality (R) | 60.20 | 22.12 | -.83*** | -.27* | -.29** | -.37*** | -- | |
| 6. Likelihood of Punishment | 39.02 | 20.09 | -.56*** | -.10 | -.50*** | -.58*** | .58*** | -- |
| 7. Level of Punishment | 34.24 | 20.52 | -.43*** | .02 | -.27* | -.73*** | .55*** | .80*** |

*Note.* Decisions: 0 = No, 1 = Yes (i.e., in favor of the action); The *SD* of the *Decisions* variable refers to variation *within* participants; All dimensions were scaled to a value range of 0 (low) – 100 (high); Morality was reverse coded, i.e., high values mean that the action was regarded as morally wrong; *** $p < .001$, ** $p < .01$, * $p < .05$

Table 11 displays the bivariate correlations between participants' personality variables, demographics, and (criminal) decisions. On average, participants showed rather high levels of self-control and low levels of sensation-seeking, psychopathy, Machiavellianism, and narcissism. Except for narcissism, which was not related to the decision outcome, all personality traits were significantly correlated with participants' decision outcomes in the expected directions ($|.28| \leq r \leq |.43|$). However, age was not associated with decisions, and gender was only weakly associated, i.e., men tended more to decide in favor of the (criminal) action.

**Table 11**

*Means (M), standard deviations (SD), and Spearman Correlations of personality traits, demographics, and positive decisions, i.e., in favor of the (criminal) action*

| | *M* | *SD* | 1. | 2. | 3. | 4. | 5. | 6. | 7. |
|---|---|---|---|---|---|---|---|---|---|
| 1. Decisions | 0.38 | 0.11 | -- | | | | | | |
| 2. Self-control | 3.29 | 0.66 | -.43*** | -- | | | | | |
| 3. Sensation-Seeking | 2.63 | 0.81 | .31*** | -.22** | -- | | | | |
| 4. Psychopathy | 2.09 | 0.62 | .35*** | -.45*** | .33*** | -- | | | |
| 5. Machiavellianism | 2.86 | 0.71 | .28*** | -.26*** | .12 | .56*** | -- | | |
| 6. Narcissism | 2.45 | 0.59 | .13 | .04 | .28*** | .30*** | .27*** | -- | |
| 7. Age | 26.65 | 5.37 | -.13 | .03 | -.02 | .02 | -.08 | .01 | -- |
| 8. Gender | 0.51 | -- | -.15* | .13 | -.26*** | -.39*** | -.27*** | -.25*** | .08 |

*Note*. Decisions: 0 = No, 1 = Yes (i.e., in favor of the action); The *SD* of the *Decisions* variable refers to variation *between* participants; All personality traits (variables #2 - #6) were scaled 1 (low) - 5 (high); Gender: 0 = male, 1 = female; *** $p < .001$, ** $p < .01$, * $p < .05$

### *Preregistered Regression Analysis*

Table 12 presents the results from the regression analysis, with separate columns for the full sample ($N = 198$) and for the reduced sample ($N = 186$; participants outside the age range 18-35 years excluded), and for unstandardized coefficients ($b$, to maintain interpretability on original scales) as well as standardized coefficients (β, to allow comparisons between scales).

Our first hypothesis was that all six appraisal dimensions have an impact on decision making, i.e., were expected to be significant (Level 1) predictors for (criminal) decisions. We found significant effects in the expected directions for *Incentive* ($b = 0.14$; 95% CI [0.11, 0.18]; $t$ (243) = 8.07, $p < .001$; β = 0.07), *Legality* ($b = 0.07$; 95% CI [0.05, 0.09]; $t$ (294) = 6.21, $p < .001$; β = 0.07), *Morality* ($b = -0.70$; 95% CI [-0.64, -0.76]: $t$ (195) = -22.37, $p < .001$; β = -0.41), and *Likelihood of Punishment* ($b = -0.25$; 95% CI [-0.20, -0.30]; $t$ (418) = -9.63, $p < .001$; β = -0.13). However, *Feasibility* did not significantly predict decision making (b = 0.02; 95% CI [−0.03, 0.06]; t (410) = 0.80, p = .425; β = 0.01).

**Table 12**

*Results of the hierarchical regression analysis for the full (N = 198) and reduced sample (n = 186)*

| | Full sample | | Reduced Sample | |
|---|---|---|---|---|
| Levels and variables | Unstandardized (*b*) | Standardized (β) | Unstandardized (b) | Standardized (β) |
| Level 1 (appraisals) | | | | |
| Intercept | 38.95*** (0.88) | 0.02 (0.02) | 39.35*** (0.92) | 0.02 (0.02) |
| Incentive | 0.14*** (0.02) | 0.07*** (0.01) | 0.15*** (0.02) | 0.07*** (0.01) |
| Feasibility | 0.02 (0.02) | 0.01 (0.01) | 0.02 (0.02) | 0.01 (0.01) |
| Legality | 0.07*** (0.01) | 0.07*** (0.01) | 0.06*** (0.01) | 0.07*** (0.01) |
| Morality | -0.70*** (0.03) | -0.41*** (0.02) | -0.71*** (0.03) | -0.42*** (0.02) |
| P(Punishment) | -0.25*** (0.03) | -0.13*** (0.01) | -0.26*** (0.03) | -0.14*** (0.01) |
| Punishment | 0.26*** (0.02) | 0.14*** (0.01) | 0.27*** (0.03) | 0.15*** (0.01) |
| Level 2 (person factors) | | | | |
| Gender | -1.48 (1.19) | -0.04 (0.03) | -1.42 (1.25) | -0.04 (0.03) |
| Age | -0.20* (0.10) | -0.03. (0.02) | -0.22 (0.14) | -0.02 (0.02) |
| Self-control | -4.17*** (0.94) | -0.07*** (0.02) | -4.03*** (0.98) | -0.07*** (0.02) |
| Psychopathy | 2.38* (1.09) | 0.04* (0.02) | 2.30* (1.13) | 0.04* (0.02) |
| Sensation-Seeking | 2.34** (0.73) | 0.05** (0.02) | 2.37** (0.75) | 0.05** (0.02) |
| Cross-Level interactions | | | | |
| Incentive x Self-control | 0.00 (0.02) | 0.00 (0.01) | 0.00 (0.03) | 0.00 (0.01) |
| Morality x Psychopathy | 0.10* (0.04) | 0.04* (0.02) | 0.10* (0.04) | 0.04* (0.02) |
| P(Punishment) x Sensation-Seeking | 0.04* (0.02) | 0.02 (0.01) | 0.04 (0.02) | 0.02 (0.01) |
| Variance components | | | | |
| Within participants | 895.58 | 0.64 | 900.12 | 0.64 |
| Intercept variance | 71.84 | 0.06 | 74.25 | 0.06 |
| Incentive variance | 0.02 | 0.00 | 0.02 | 0.00 |
| Feasibility variance | 0.01 | 0.00 | 0.00 | 0.00 |
| Legality variance | 0.01 | 0.01 | 0.01 | 0.01 |
| Morality variance | 0.15 | 0.05 | 0.15 | 0.05 |
| P(Punishment) variance | 0.02 | 0.01 | 0.02 | 0.01 |
| Punishment variance | 0.02 | 0.00 | 0.02 | 0.00 |
| Additional Information | | | | |
| Number of groups | 198 | 198 | 186 | 186 |
| Number of observations | 15806 | 15806 | 14847 | 14847 |

*Note.* Values in parentheses represent standard error; *t*-statistics were computed as the ratio of each regression coefficient divided by its standard error; *** $p < .001$, ** $p < .01$, * $p < .05$.

Furthermore, *Level of Punishment* was a significant predictor, however, its coefficient was positive ($b = 0.26$), which means that a higher *Level of Punishment* lead to a higher probability of committing the (criminal) behavior, which was contrary to our hypothesis. In contrast to that, the zero-order correlation of *Level of Punishment* with (criminal) decisions was negative ($r = -.43$, Table 1), which was in line with our hypothesis. Presumably, the negative effect in the regression model resulted from a high collinearity between the two

dimensions *Likelihood* and *Level of Punishment* ($r$ = .80), so the parameter estimation could have been biased (Mason & Perreault, 1991).

Our second hypothesis was that interindividual differences in the relevance (i.e., weight) of the dimensions for decision making can be explained by differences in stable person factors. In order to test whether person factors accounted for some of the variation, we specified three cross-level interactions (*Morality* x psychopathy, *Likelihood of Punishment* x sensation-seeking, *Incentive* x self-control). Table 12 shows that, generally, variance components were rather small, i.e., participants only differed to a small extent regarding the weighting of the dimensions, apart from *Morality*, where variance was considerably higher. As predicted, we found significant cross-level interactions between *Morality* and psychopathy (i.e., participants scoring high in psychopathy put less weight on the *Morality* dimension; $b$ = 0.10; 95% CI [0.02, 0.18]; $t$ (203) = 2.50, $p$ = .013; β = 0.04), as well as between *Likelihood of Punishment* and sensation-seeking (i.e., participants scoring high in sensation-seeking put less weight on the *Likelihood of Punishment*; $b$ = 0.04; 95% CI [0.00, 0.08]; $t$ (277) = 1.98, $p$ = .049; β = 0.02). However, the interaction between *Likelihood of Punishment* and sensation-seeking was very small and was no longer significant when the analysis was conducted with the reduced sample ($t$ (259) = 1.75, $p$ = .081). Contrary to our hypothesis, there was no interaction between *Incentive* and self-control ($b$ = 0.00; 95% CI [-0.05, 0.05]; $t$ (308) = 0.14, $p$ = .892; β = 0.00).

Additionally, we found that self-control, sensation-seeking, and psychopathy also affected decisions more directly, i.e., without consideration of the appraisal dimensions. The high variance estimation of the intercept ($\sigma^2$ = 71.84) indicates that participants also differed regardless of their weighting of the appraisal dimensions. Some of this variation was accounted for by self-control ($b$ = −4.17; 95% CI [−2.33, −6.01]; $t$ (225) = −4.45, $p$ < .001; β = −0.07), sensation-seeking ($b$ = 2.34; 95% CI [0.92, 3.76]; $t$ (224) = 3.23, $p$ = .001; β = 0.05), and psychopathy ($b$ = 2.38; 95% CI [0.24, 4.51]; $t$ (227) = 2.18, $p$ = .030; β = 0.04). Regarding demographic variables, we only found a small significant effect for age, and only when analyzing the full sample ($b$ = -0.20; $t$ (222) = −1.99, $p$ = .048), and no effect for gender.

To comprehensively evaluate the fit of our model, we computed several fit indices (AIC, BIC, adjusted $R^2$) and compared our model to different model alternatives (non-hierarchical, random intercept, random slopes, Level 1 only). Results showed that the preregistered model

provided the best fit to the data compared to alternative models, and that the model benefited from the consideration of individual differences and their interaction with appraisal dimensions. The whole model fit evaluation can be found in Appendix I.

### *Exploratory Regression Analysis*

In addition to testing the three preregistered cross-level interaction hypotheses, we also explored whether other person factors interacted with the weighting of the appraisal dimensions. First, we tested whether participants low in self-control put less weight on the *Likelihood of Punishment* dimension because similar effects were found by Nagin and Paternoster (1993). However, we did not find such an interaction effect in the present study ($b$ = 0.02; 95% CI [-0.03, 0.07]; $t$ (301) = 0.71, $p$ = .481; β = 0.01). Furthermore, we tested whether psychopathy and sensation-seeking interacted with the *Legality* dimension because previous studies (Studies 1, 2 and 4) suggested that participants high in psychopathy/ sensation-seeking might not care whether an action is illegal or may even find illegal actions particularly appealing. In fact, we found significant cross-level interactions between psychopathy and *Legality* ($b$ = -0.04; 95% CI [-0.01, -0.07]; $t$ (218) = -2.88, $p$ = .004; β = -0.03) and between sensation-seeking and *Legality* ($b$ = -0.03; 95% CI [-0.01, -0.05]; $t$ (204) = -2.84, $p$ = .005; β = -0.03). To participants scoring high in psychopathy and sensation-seeking it was less important if an action was illegal when they made their decisions. As the regression coefficient for *Legality* was rather small ($b$/β = .07) and the coefficients for the cross-level interactions (*Legality* x psychopathy, *Legality* x sensation-seeking) were comparatively large (-.03 ≥ $b$/β ≥ -.04), the coefficient could even become negative for people who score very high on psychopathy and sensation-seeking. That means to high scorers, whether an action is illegal does play a role, however, illegal actions are considered particularly appealing and the execution of these actions will become more likely.

## Discussion

In the present vignette study, we used a comprehensive regression model to investigate the appraisal dimensions' influence on decision making and how individual difference factors impact the weighting of the appraisal dimensions in the process of decision making. The results showed that the dimensions *Incentive*, *Legality*, *Morality,* and *Likelihood of Punishment* significantly predicted decision making, with *Morality* showing the strongest

impact. *Level of Punishment* had a significant impact on decisions as well. However, the negative coefficient in the regression indicated that higher levels of punishment increased the chance of a (hypothetical) behavior execution. As the zero-order correlation between *Level of Punishment* and (criminal) decisions was positive, and there was a high collinearity between the two predictors *Likelihood* and *Level of Punishment*, we regard it as a statistical artefact that the sign of the beta coefficient of *Level of Punishment* pointed in the wrong direction. Moreover, *Feasibility* did not predict decision making at all.

Regarding the interaction of person factors with appraisal dimensions, the results showed that participants differed in the way their decisions were influenced by the dimensions. We found that participants high in psychopathy put less weight on the *Morality* and *Legality* dimensions and participants high in sensation-seeking put less weight on the *Likelihood of Punishment* (although this effect was rather unstable) and on the *Legality* dimension. Our hypothesis that participants low in self-control weighted the *Incentive* more strongly was not supported.

In an exploratory analysis, we tested whether self-control interacted with *Likelihood of Punishment* because Nagin and Paternoster (1993) found that participants low in self-control perceived the potential costs of crime as less aversive. However, Piquero and Tibbetts (1996) did not find any effect of self-control on perceived sanctions in their scenario study, and we did not find this interaction effect in the present study.

However, psychopathy interacted with the *Legality* dimension in that participants with higher psychopathy scores did not put much weight on whether an action was legal or illegal. The same was found for sensation-seeking. In our qualitative interview study with former offenders (Study 4), many of the participants stated that they found the fact that an act is illegal particularly appealing. In other words, for them the illegality of an act was not an obstacle, but rather an incentive. Evidence of this effect was also found in the present study. Based on the regression weights, it can be estimated that for people who score very high on psychopathy and sensation-seeking, the relevance of the *Legality* dimension increases, however, the effect tilts in the opposite direction. That means the appraisal that an action is illegal will lead to an increase in the probability that the action will be carried out. As participants in the present study generally showed rather low levels of psychopathy and sensation-seeking, it would be interesting to investigate this in a sample of individuals high in psychopathy and sensation-seeking.

In our theoretical model, we conceptualized that people differ regarding the weighting of the appraisal dimensions, but also regarding how they appraise actions in specific situations. The concept of weightings could easily be transferred to the mathematical model, as the regression weights reflect the importance of the appraisal dimensions. However, the interaction effects could also result from a different appraisal of the action (i.e., different setpoints). For methodological reasons, we did not ask the main study participants themselves about the appraisals but used the averaged appraisals from pre-study participants as predictors. When a participant from the main study decided in favor of an action that was appraised as very problematic with regards to a certain dimension by the pre-study participants, this decision could either be caused by a different weighting of the dimensions or by a different appraisal, i.e., by a shifted setpoint of this participant. More specifically, it is unclear whether people high in psychopathy appraise something as morally wrong but still execute it because they do not care, or if they appraise it as morally more acceptable than others do and execute it for this reason. Blair (1995) found that participants with high psychopathy scores ($M$ = 31.6, $SD$ = 2.1 out of 40 on the Psychopathy Checklist by Hare, 1980) were unable to make a distinction between moral transgressions and conventional rule transgressions, i.e., they judged conventional transgressions as equally serious and permissible as moral transgressions. This finding supports the idea that participants high in psychopathy have a shifted moral setpoint and therefore appraise certain actions as morally less problematic than other participants appraise the actions.

In order to further investigate the influence of participants' individual setpoints, appraisals and decisions need to be measured in one sample. However, there should be a temporal delay between the two tasks to avoid a mutual contamination of the appraisal and decision ratings.

### *Limitations*

As in the other vignette studies, one fundamental limitation was that we have not measured actual behavior but hypothetical behavior, or behavioral intentions. Nevertheless, the vignette paradigm was chosen because vignettes offer a unique opportunity to assess a broad range of situations in a within participant design.

Another limitation was the high collinearity between *Likelihood* and *Level of Punishment*, which may have distorted the effects. In Studies 1-3, we found that participants had

difficulties rating the dimensions independently from each other. Therefore, in the present study, we have addressed this problem with very detailed instructions and a practice trial. When constructing the vignettes, we also attempted to make sure that the severity and likelihood of a punishment were not too strongly confounded with each other (severe punishments were often unlikely and minor punishments were often likely to occur). However, participants still rated these dimensions similarly. This indicates shortcomings in the data quality of the pre-study, which was conducted via the Prolific Academic participant platform. On closer inspection, we noticed that the average total processing time that participants needed to read the instructions, perform the exercise trial, and rate all 42 vignettes regarding six dimensions was rather short ($M$ = 29.8 minutes, $SD$ = 6.5). Furthermore, only one participant informed us about a programming error in one of the vignettes. Regarding future studies, it should be ascertained that these two dimensions in fact exhibit enough independent variation across vignettes, and the vignettes should be rated by a group of trained experts.

Furthermore, there was no effect of *Feasibility*, which was presumably also due to a lack of variation of this dimension, and a deficient construction of the vignettes. To prevent the majority of actions from not even being considered, we limited feasibility in only a small number of vignettes. Unfortunately, participants were confused by some of these vignettes. For example, in one vignette participants were asked whether they would secretly read messages on someone else's mobile phone, although they did not know the phone's unlock code. In this scenario it was virtually impossible to carry out the action, thus, there was no real decision to make. Some (main study) participants reported that they considered the information about limited feasibility of the action to be a construction or programming error.

Finally, there was high variation in the intercept which suggests that there might be other appraisals or processes that were not addressed in our study. This illustrates that, although it was possible to translate some person factors into appraisals, it is overly simplistic or optimist to expect that the complexity of criminal behavior can be fully explained in terms of appraisals.

### *Conclusion*

The present study provided evidence that largely confirmed the predictions of the appraisal model. However, although most of the appraisal dimensions impacted criminal decision making, *Feasibility* did not. Further studies should investigate the potential influence of *Feasibility*, as well as the question whether *Likelihood* and *Level of Punishment* constitute unique appraisal dimensions. In future studies, it could also be enlightening to test the vignette paradigm in a forensic sample with more variation regarding criminogenic personality factors. This way, it could be examined whether the appraisal patterns also apply to more severe crimes, which has so far only been explored in a qualitative interview study (Study 4). Furthermore, besides rather stable personality aspects, it remains to be investigated in more detail how dynamic factors such as arousal or intoxication affect appraisals and, as a consequence, criminal decision making. Another future direction could be development and testing of more complex models, e.g., cut-off models, and exploring whether appraisals are processed in parallel or serially.

## General Conclusion and Outlook

In this thesis, a new model of criminal decision making was presented that introduces the concept of appraisals to the domain of legal psychology. The model combines individual difference (personality) theories (e.g., Gottfredson & Hirschi, 1990) with theories focusing on perceptions of specific situations, (e.g., Cohen & Felson, 1979; Cornish & Clarke, 1986, 2011). The interaction results from the vignette studies support the idea that personality factors can be translated into specific appraisal patterns that might provide a causal link between established risk factors and actual criminal behavior.

Despite the theoretical and empirical limitations described above, the approach offers potential in several domains. On a theoretical level, our approach builds on the concept of validity by Borsboom et al. (2004). The authors argued that a test is valid if variations in the attribute *causally* produce variation in the measurement outcomes and that research must be directed at these causal processes that convey correlative effects. We have followed this approach by mapping correlations between person factors and crime in terms of appraisal patterns that might be underlying the correlation. In his call to pay substantially more attention to choice in criminological theory and research, also Nagin (2007) argued that

understanding the developmental course of criminal decision making is crucial to understanding issues such as the emergence of crime from childhood problem behaviors, the chronic choice of crime, and desistance from crime.

Understanding the underlying psychological mechanisms is also essential when it comes to treating and changing criminal behavior. It might be possible to determine individual offenders' current appraisal profiles, or to identify the appraisals that were relevant to their criminal history. If therapists understand how personality traits operate in the very moment of decision making, they could simulate these situations and help clients change their specific appraisal patterns.

Identifying appraisal patterns and their relation to recidivism could also help to improve risk assessment. In recent years, studies have shown that spontaneous changes of goals and values play an important role for desistance from crime (Giordano et al., 2002; Laub & Sampson, 2001; Maruna et al., 2004). Actuarial risk assessment tools and also measures of criminogenic personality traits are mainly based on past behavior (Hanson & Thornton, 1999; Rettenberger et al., 2010). These approaches are reaching their limits in the context of desistance because they are unable to account for these changes in goals and values (Maruna, 2012). We therefore hope that future studies will provide more insights into the interplay among trait factors, appraisals, and criminal decisions, and thus help to predict recidivism.

The model proposed in this thesis puts emphasis on appraisal processes in criminal decision making that have been largely neglected. First evidence for the validity of the model was presented, and methodological problems were discussed. Although further research is needed, the results indicate that the appraisal model can be a useful addition to the existing theoretical landscape and might contribute to a better understanding of why people sometimes choose to commit criminal actions.

## References


Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. (2013). Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *Journal of Management*, *39*(6), 1490–1528. https://doi.org/10.1177/0149206313478188

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, *80*(1), 123–129. https://doi.org/10.1016/S0165-1765(03)00032-6

Akers, R. L. (1991). Self-control as a general theory of crime. *Journal of Quantitative Criminology*, *7*(2), 201–211.

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct*. Routledge.

Ariely, D., & Loewenstein, G. (2006). The heat of the moment: the effect of sexual arousal on sexual decision making. *Journal of Behavioral Decision Making*, *19*(2), 87–98. https://doi.org/10.1002/bdm.501

Asencio, E. K., & Burke, P. J. (2011). Does incarceration change the criminal identity? A synthesis of labeling and identity theory perspectives on identity change. *Sociological Perspectives*, *54*(2), 163–182. https://doi.org/10.1525%2Fsop.2011.54.2.163

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, *45*(2), 527–535. https://doi.org/10.3758/s13428-012-0265-2

Azizli, N., Atkinson, B. E., Baughman, H. M., Chin, K., Vernon, P. A., Harris, E., & Veselka, L. (2016). Lies and crimes: Dark Triad, misconduct, and high-stakes deception. *Personality and Individual Differences*, *89*, 34–39. https://doi.org/10.1016/j.paid.2015.09.034

Bachman, R., Paternoster, R., & Ward, S. (1992). The rationality of sexual offending: Testing a deterrence/rational choice conception of sexual assault. *Law & Society Review*, *26*(2), 343–372. https://doi.org/10.2307/3053901

Baddeley, A., Banse, R., Huang, Y.-M., & Page, M. (2012). Working memory and emotion: Detecting the hedonic detector. *Journal of Cognitive Psychology*, *24*(1), 6–16. https://doi.org/10.1080/20445911.2011.613820

Beier, K. M. (1998). Differential typology and prognosis for dissexual behavior–a follow-up study of previously expert-appraised child molesters. *International Journal of Legal Medicine*, *111*, 133–141. https://doi.org/10.1007/s004140050133

Bertrams, A., & Dickhäuser, O. (2009). Messung dispositioneller Selbstkontroll-Kapazität. *Diagnostica*, *55*(1), 2–10. https://doi.org/10.1026/0012-1924.55.1.2

Blair, K. S., Morton, J., Leonard, A., & Blair, R. J.R. (2006). Impaired decision-making on the basis of both reward and punishment information in individuals with psychopathy. *Personality and Individual Differences*, *41*(1), 155–165. https://doi.org/10.1016/j.paid.2005.11.031

Blair, R. J. R. (1995). A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, *57*(1), 1–29. https://doi.org/10.1016/0010-0277(95)00676-P

Blair, R. J. R., Mitchell, D. G.V., Leonard, A., Budhani, S., Peschardt, K. S., & Newman, C. (2004). Passive avoidance learning in individuals with psychopathy: Modulation by reward but not by punishment. *Personality and Individual Differences*, *37*(6), 1179–1192. https://doi.org/10.1016/j.paid.2003.12.001

Blickle, G., Schlegel, A., Fassbender, P., & Klein, U. (2006). Some personality correlates of business white‐collar crime. *Applied Psychology*, *55*(2), 220–233. https://doi.org/10.1111/j.1464-0597.2006.00226.x

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Bouffard, J. A. (2002). The influence of emotion on rational decision making in sexual aggression. *Journal of Criminal Justice*, *30*(2), 121–134. https://doi.org/10.1016/S0047-2352(01)00130-1

Bouffard, J. A. (2011). 'In the heat of the moment': mediating versus moderating relationships between sexual arousal and perceived sanctions. *Journal of Crime and Justice*, *34*(1), 24–44. https://doi.org/10.1080/0735648X.2011.554745

Broidy, L., & Agnew, R. (1997). Gender and crime: A general strain theory perspective. *Journal of Research in Crime and Delinquency*, *34*(3), 275–306. https://doi.org/10.1177/0022427897034003001

Carmichael, S., & Piquero, A. R. (2004). Sanctions, perceived anger, and criminal offending, *20*(4), 371–393. https://doi.org/10.1007/s10940-004-5869-y

Cleckley, H. (1988). *The mask of sanity - An attempt to clarify some issues about the so-called psychopathic personality* (Fifth Edition: Private printing for non-profit educational use). https://pdfs.semanticscholar.org/1bc9/2308d567a435174d427796a410eb1b5bf211.pdf

Cocheo, S. (1997). The bank robber, the quote, and the final irony. *ABA Banking Journal*, *89*(3), 71.

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, *44*(4), 588–608. DOI: 10.2307/2094589

Cole, C. A. (1989). Deterrence and consumer fraud. *Journal of Retailing*, *65*(1), 107–120.

Cornish, D. B., & Clarke, R. V. (Eds.). (1986). *The reasoning criminal: Rational choice perspectives on offending.* Springer.

Cornish, D. B., & Clarke, R. V. (2011). The rational choice perspective. In R. Wortley & M. Townsley (Eds.), *Environmental criminology and crime analysis* (pp. 21–44). Routledge.

De Haan, W., & Vos, J. (2003). A crying shame: The over-rationalized conception of man in the rational choice perspective. *Theoretical Criminology*, *7*(1), 29–54. https://doi.org/10.1177/1362480603007001199

De Jong, P. J., van Overveld, M., & Borg, C. (2013). Giving in to arousal or staying stuck in disgust? Disgust-based mechanisms in sex and sexual dysfunction. *Journal of Sex Research*, *50*(3-4), 247–262. https://doi.org/10.1080/00224499.2012.746280

De Li, S. (2004). The impacts of self-control and social bonds on juvenile delinquency in a national sample of midadolescents. *Deviant Behavior*, *25*(4), 351–373. https://doi.org/10.1080/01639620490441236

DeLisi, M. (2009). Psychopathy is the unified theory of crime. *Youth Violence and Juvenile Justice*, *7*(3), 256–273. https://doi.org/10.1177/1541204009333834

Diedenhofen, B., & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *PloS One*, *10*(4), 1-12. https://doi.org/10.1371/journal.pone.0121945

Eickmeier, K., Hoffmann, L., & Banse, R. (2019). The 5-Factor Disgust Scale: Development and validation of a comprehensive measure of disgust Propensity. *European Journal of Psychological Assessment*, *35*(3), 403–413. https://doi.org/10.1027/1015-5759/a000401

Eisenbarth, H., Lilienfeld, S. O., & Yarkoni, T. (2015). Using a genetic algorithm to abbreviate the Psychopathic Personality Inventory–Revised (PPI-R). *Psychological Assessment*, *27*(1), 194. https://doi.org/10.1037/pas0000032

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572–595). Oxford University Press.

Farrington, D. P. (1986). Age and crime. In *Crime and justice: An annual review of research, vol 7.* (pp. 189–250). University of Chicago Press.

Fedina, L., Holmes, J. L., & Backes, B. L. (2018). Campus sexual assault: A systematic review of prevalence research from 2000 to 2015. *Trauma, Violence, & Abuse*, *19*(1), 76–93. https://doi.org/10.1177/1524838016631129

Finkelhor, D., Shattuck, A., Turner, H. A., & Hamby, S. L. (2014). The lifetime prevalence of child sexual abuse and sexual assault assessed in late adolescence. *Journal of Adolescent Health*, *55*(3), 329–333. https://doi.org/10.1016/j.jadohealth.2013.12.026

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Addison-Wesley.

Ford, C. A., & Donis, F. J. (1996). The relationship between age and gender in workers' attitudes toward sexual harassment. *The Journal of Psychology*, *130*(6), 627–633. https://doi.org/10.1080/00223980.1996.9915036

Giordano, P. C., Cernkovich, S. A., & Rudolph, J. L. (2002). Gender, crime, and desistance: Toward a theory of cognitive transformation. *American Journal of Sociology*, *107*(4), 990–1064.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford University Press.

Grasmick, H. G., Tittle, C. R., Bursik Jr, R. J., & Arneklev, B. J. (1993). Testing the core empirical implications of Gottfredson and Hirschi's general theory of crime. *Journal of*

*Research in Crime and Delinquency*, *30*(1), 5–29. https://doi.org/10.1177/0022427893030001002

Green, D. E. (1989). Measures of illegal behavior in individual-level deterrence research. *Journal of Research in Crime and Delinquency*, *26*(3), 253–275. https://doi.org/10.1177/0022427889026003004

Hammond, M. D., Milojev, P., Huang, Y., & Sibley, C. G. (2018). Benevolent sexism and hostile sexism across the ages. *Social Psychological and Personality Science*, *9*(7), 863–874. https://doi.org/10.1177/1948550617727588

Hanson, R. K., & Thornton, D. (1999). *Static 99: Improving actuarial risk assessments for sex offenders*. Solicitor General Canada.

Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and Individual Differences*, *1*(2), 111–119. https://doi.org/10.1016/0191-8869(80)90028-8

Hare, R. D., & Neumann, C. S. (2009). Psychopathy: Assessment and forensic implications. *Canadian Journal of Psychiatry*, *54*(12), 791–802. https://doi.org/10.1177/070674370905401202

Hart, S. D., & Hare, R. D. (1997). Psychopathy: Assessment and association with criminal conduct. In D. M. Stoff, J. Breiling, & J. D. Maser (Eds.), *Handbook of antisocial behavior* (pp. 22–35). Wiley.

Hepper, E. G., Hart, C. M., Meek, R., Cisek, S., & Sedikides, C. (2014). Narcissism and empathy in young offenders and non‐offenders. *European Journal of Personality*, *28*(2), 201–210. https://doi.org/10.1002/per.1939

Hirtenlehner, H., & Kunz, F. (2016). The interaction between self-control and morality in crime causation among older adults. *European Journal of Criminology*, *13*(3), 393–409. https://doi.org/10.1177/1477370815623567

Imhoff, R., & Schmidt, A. F. (2014). Sexual disinhibition under sexual arousal: Evidence for domain specificity in men and women. *Archives of Sexual Behavior*, *43*(6), 1123–1136. https://doi.org/10.1007/s10508-014-0329-8

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion-is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological Science*, *21*(11), 1686–1693. https://doi.org/10.1177/0956797610384745

Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A brief measure of dark personality traits. *Assessment*, *21*(1), 28–41. https://doi.org/10.1177/1073191113514105

Kafka, M. P. (1997). Hypersexual desire in males: An operational definition and clinical implications for males with paraphilias and paraphilia-related disorders. *Archives of Sexual Behavior*, *26*(5), 505–526. https://doi.org/10.1023/A:1024507922470

Kim, M.-S., & Hunter, J. E. (1993). Relationships among attitudes, behavioral intentions, and behavior: A meta-analysis of past research, part 2. *Communication Research*, *20*(3), 331–364. https://doi.org/10.1177/009365093020003001

Kinsey, A. C., Pomeroy, W. B., & Martin, C. E. (1948). *Sexual behavior in the human male*. W. B. Saunders.

Klepper, S., & Nagin, D. (1989). The deterrent effect of perceived certainty and severity of punishment revisited. *Criminology*, *27*(4), 721–746. https://doi.org/10.1111/j.1745-9125.1989.tb01052.x

Kraut, R. E. (1976). Deterrent and Definitional Influences on Shoplifting. *Social Problems*, *23*(3), 358–368. https://doi.org/10.2307/799781

Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* Meredith Corporation.

Laub, J. H., & Sampson, R. J. (2001). Understanding desistance from crime. *Crime and Justice*, *28*, 1–69. https://doi.org/10.1086/652208

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Loewenstein, G., Nagin, D., & Paternoster, R. (1997). The effect of sexual arousal on expectations of sexual forcefulness. *Journal of Research in Crime and Delinquency*, *34*(4), 443–473. https://doi.org/10.1177/0022427897034004003

Lösel, F., & Schmucker, M. (2008). Kriminalitätstheorien: Theories of crime. In R. Volbert & M. Steller (Eds.), *Handbuch der Rechtspsychologie* (pp. 15–27). Hogrefe.

Malesza, M., Ostaszewski, P., Büchner, S., & Kaczmarek, M. C. (2019). The adaptation of the Short Dark Triad personality measure – Psychometric properties of a German sample. *Current Psychology*, *38*(3), 855–864. https://doi.org/10.1007/s12144-017-9662-0

Maruna, S. (2012). Elements of successful desistance signaling. *Criminology & Public Policy*, *11*, 73–86. https://doi.org/10.1111/j.1745-9133.2012.00789x

Maruna, S., Porter, L., & Carvalho, I. (2004). The Liverpool desistance study and probation practice: Opening the dialogue. *Probation Journal*, *51*(3), 221–232. https://doi.org/10.1177/0264550504045899

Mason, C. H., & Perreault, W. D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research*, *28*(3), 268–280. https://doi.org/10.1177/002224379102800302

Maxwell, S. R., & Gray, M. K. (2007). Deterrence: Testing the Effects of Perceived Sanction Certainty on Probation Violations. *Sociological Inquiry*, *70*(2), 117–136. https://doi.org/10.1111/j.1475-682X.2000.tb00901.x

Mayring, P. (2000). Qualitative content analysis. *Forum Qualitative Social Research*, *1*(2), 1–10. http://dx.doi.org/10.17169/fqs-1.2.1089

Miller, J. D., & Lynam, D. (2001). Structural models of personality and their relation to antisocial behavior: A meta‐analytic review. *Criminology*, *39*(4), 765–798. https://doi.org/10.1111/j.1745-9125.2001.tb00940.x

Mills, J. F., Loza, W., & Kroner, D. G. (2003). Predictive validity despite social desirability: Evidence for the robustness of self‐report among offenders. *Criminal Behaviour and Mental Health*, *13*(2), 140–150. https://doi.org/10.1002/cbm.536

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, *100*(4), 674–701. https://doi.org/10.1037/0033-295X.100.4.674

Müller, U., & Schröttle, M. (2004). *Lebenssituation, Sicherheit und Gesundheit von Frauen in Deutschland. Eine repräsentative Untersuchung zu Gewalt gegen Frauen in Deutschland. Im Auftrag des Bundesministeriums für Familie, Senioren, Frauen und Jugend.* https://www.bmfsfj.de/blob/84328/0c83aab6e685eeddc01712109bcb02b0/langfassung-studie-frauen-teil-eins-data.pdf

Murray, G. F., & Erickson, P. G. (1987). Cross-sectional versus longitudinal research: An empirical comparison of projected and subsequent criminality. *Social Science Research*, *16*(2), 107–118. https://doi.org/10.1016/0049-089X(87)90011-1

Nagin, D. S. (2007). Moving choice to center stage in criminological research and theory: The american society of criminology 2006 Sutherland address. *Criminology*, *45*(2), 259-272.

Nagin, D. S., & Paternoster, R. (1993). Enduring individual differences and rational choice theories of crime. *Law & Society Review*, *27*(3), 467–496. https://doi.org/10.2307/3054102

Nagin, D. S., & Paternoster, R. (1994). Personal capital and social control: The deterrence implications of a theory of individual differences in criminal offending. *Criminology*, *32*(4), 581–606. https://doi.org/10.1111/j.1745-9125.1994.tb01166.x

Nagin, D. S., & Pogarsky, G. (2001). Integrating celerity, impulsivity, and extralegal sanction threats into a model of general deterrence: Theory and evidence. *Criminology*, *39*(4), 865–892. https://doi.org/10.1111/j.1745-9125.2001.tb00943.x

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Newburn, T., & Stanko, E. A. (Eds.). (1994). *Men, masculinities and crime - Just boys doing business?* Routledge.

Newman, J. P., & Kosson, D. S. (1986). Passive avoidance learning in psychopathic and nonpsychopathic offenders. *Journal of Abnormal Psychology*, *95*(3), 252. https://doi.org/10.1037/0021-843X.95.3.252

Oliver, B. E. (2007). Preventing female-perpetrated sexual abuse. *Trauma, Violence, & Abuse*, *8*(1), 19–32. https://doi.org/10.1177/1524838006296747

Paternoster, R., & Simpson, S. (1996). Sanction threats and appeals to morality: Testing a rational choice model of corporate crime. *Law & Society Review*, *30*(3), 549–583. https://doi.org/10.2307/3054128

Paulhus, D. L., Hemphill, J. D., & Hare, R. D. (2015). *Self-Report Psychopathy scale: Version III*. Multi-Health Systems.

Penke, L. (2011). The revised Sociosexual Orientation Inventory. In T. D. Fisher, C. M. Davis, & W. L. Yarber (Eds.), *Handbook of sexuality-related measures* (pp. 622–625). Routledge. https://doi.org/10.1037/0022-3514.95.5.1113

Penke, L., & Asendorpf, J. B. (2008). Beyond global sociosexual orientations: a more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *Journal of Personality and Social Psychology*, *95*(5), 1113–1135. https://doi.org/10.1037/0022-3514.95.5.1113

Piquero, A., & Tibbetts, S. (1996). Specifying the direct and indirect effects of low self-control and situational factors in offenders' decision making: Toward a more complete model of rational offending. *Justice Quarterly*, *13*(3), 481–510. https://doi.org/10.1080/07418829600093061

Pogarsky, G. (2002). Identifying “deterrable” offenders: Implications for research on deterrence. *Justice Quarterly*, *19*(3), 431–452. https://doi.org/10.1080/07418820200095301

Pogarsky, G. (2004). Projected offending and contemporaneous rule‐violation: Implications for heterotypic continuity. *Criminology*, *42*(1), 111–136. https://doi.org/10.1111/j.1745-9125.2004.tb00515.x

Pratt, T. C. (2008). Rational choice theory, crime control policy, and criminological relevance. *Criminology & Public Policy*, *7*(1), 43–52.

Pratt, T. C., & Cullen, F. T. (2000). The empirical status of Gottfredson and Hirschi's general theory of crime: A meta‐analysis. *Criminology*, *38*(3), 931–964. https://doi.org/10.1111/j.1745-9125.2000.tb00911.x

Pratt, T. C., Cullen, F. T., Blevins, K. R., Daigle, L. E., & Madensen, T. D. (2006). The empirical status of deterrence theory: A meta-analysis. In F. T. Cullen & K. R. Blevins (Eds.), *Taking stock: The status of criminological theory* (pp. 367–396). Transaction Publishers.

Rammstedt, B., & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K). *Diagnostica*, *51*(4), 195–206. https://doi.org/10.1026/0012-1924.51.4.195

Rettenberger, M., Matthes, A., Boer, D. P., & Eher, R. (2010). Prospective actuarial risk assessment: A comparison of five risk assessment instruments in different sexual offender subtypes. *International Journal of Offender Therapy and Comparative Criminology*, *54*(2), 169–186. https://doi.org/10.1177/0306624X08328755

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology*, *84*(5), 754. https://doi.org/10.1037/0021-9010.84.5.754

Roberti, J. W. (2004). A review of behavioral and biological correlates of sensation seeking. *Journal of Research in Personality*, *38*(3), 256–279. https://doi.org/10.1016/S0092-6566(03)00067-9

Roth, M., & Hammelstein, P. (2012). The Need Inventory of Sensation Seeking (NISS). *European Journal of Psychological Assessment*, *28*(1), 11–18. https://doi.org/10.1027/1015-5759/a000085

Roth, M., Schumacher, J., & Brähler, E. (2005). Sensation seeking in the community: Sex, age and sociodemographic comparisons on a representative German population sample. *Personality and Individual Differences*, *39*(7), 1261–1271. https://doi.org/10.1016/j.paid.2005.05.003

Scherer, K. R. (1997). Profiles of emotion-antecedent appraisal: Testing theoretical predictions across cultures. *Cognition & Emotion*, *11*(2), 113–150. https://doi.org/10.1080/026999397379962

Scherer, K. R. (2005). Unconscious processes in emotion: The bulk of the iceberg. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 312–334). Guilford Press.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition & Emotion*, *23*(7), 1307–1351. https://doi.org/10.1080/02699930902928969

Schoepfer, A., & Piquero, A. R. (2006). Self-control, moral beliefs, and criminal activity. *Deviant Behavior*, *27*(1), 51–71. https://doi.org/10.1080/016396290968326

Sheeran, P., & Webb, T. L. (2016). The intention-behavior gap. *Social and Personality Psychology Compass*, *10*(9), 503–518. https://doi.org/10.1111/spc3.12265

Silberman, M. (1976). Toward a theory of criminal deterrence. *American Sociological Review*, *41*(3), 442–461. https://doi.org/10.2307/2094253

Stevenson, R. J., Case, T. I., & Oaten, M. J. (2011). Effect of self-reported sexual arousal on responses to sex-related and non-sex-related disgust cues. *Archives of Sexual Behavior*, *40*(1), 79–85. https://doi.org/10.1007/s10508-009-9529-z

Svensson, R., Pauwels, L., & Weerman, F. M. (2010). Does the effect of self-control on adolescent offending vary by level of morality? A test in three countries. *Criminal Justice and Behavior*, *37*(6), 732–743. https://doi.org/10.1177/0093854810366542

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self‐control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, *72*(2), 271–324. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Thurman, Q., Jackson, S., & Zhao, J. (1993). Drunk-driving research and innovation: A factorial survey study of decisions to drink and drive. *Social Science Research*, *22*(3), 245–264. https://doi.org/10.1006/ssre.1993.1012

Tittle, C. R., Antonaccio, O., Botchkovar, E., & Kranidioti, M. (2010). Expected utility, self-control, morality, and criminal probability. *Social Science Research*, *39*(6), 1029–1046. https://doi.org/10.1016/j.ssresearch.2010.08.007

Tittle, C. R., & Rowe, A. R. (1974). Certainty of arrest and crime rates: A further test of the deterrence hypothesis. *Social Forces*, *52*(4), 455–462. https://doi.org/10.1093/sf/52.4.455

Trevethan, S. D., & Walker, L. J. (1989). Hypothetical versus real-life moral reasoning among psychopathic and delinquent youth. *Development and Psychopathology*, *1*(2), 91–103. https://doi.org/10.1017/S0954579400000286

Umbach, R., Berryessa, C. M., & Raine, A. (2015). Brain imaging research on psychopathy: Implications for punishment, prediction, and treatment in youth and adults. *Journal of Criminal Justice*, *43*(4), 295–306. https://doi.org/10.1016/j.jcrimjus.2015.04.003

van Gelder, J.-L. (2013). Beyond rational choice: The hot/cool perspective of criminal decision making. *Psychology, Crime & Law*, *19*(9), 745–763. https://doi.org/10.1080/1068316X.2012.660153

van Gelder, J.-L., Luciano, E. C., Weulen Kranenbarg, M., & Hershfield, H. E. (2015). Friends with my future self: Longitudinal vividness intervention reduces delinquency. *Criminology*, *53*(2), 158–179. https://doi.org/10.1111/1745-9125.12064

van Gelder, J.-L., & Vries, R. E. de (2012). Traits and states: Integrating personality and affect into a model of criminal decision making. *Criminology*, *50*(3), 637–671. https://doi.org/10.1111/j.1745-9125.2012.00276.x

van Gelder, J.-L., & Vries, R. E. de (2014). Rational misbehavior? Evaluating an integrated dual-process model of criminal decision making. *Journal of Quantitative Criminology*, *30*(1), 1–27. https://doi.org/10.1007/s10940-012-9192-8

Vazsonyi, A. T., Mikuška, J., & Kelley, E. L. (2017). It's time: A meta-analysis on the self-control-deviance link. *Journal of Criminal Justice*, *48*, 48–63. https://doi.org/10.1016/j.jcrimjus.2016.10.001

Waldo, G. P., & Chiricos, T. G. (1972). Perceived penal sanction and self-reported criminality: A neglected approach to deterrence research. *Social Problems*, *19*(4), 522–540. https://doi.org/10.2307/799929

Waubert de Puiseau, B., Glöckner, A., & Towfigh, E. V. (2019). Integrating theories of law obedience: How utility-theoretic factors, legitimacy, and lack of self-control influence decisions to commit low-level crimes. *Judgement and Decision Making*, *14*(3), 318-334.

Wikström, P.-O. H. (2004). Crime as alternative. Towards a cross-level situational action theory of crime causation. In J. McCord (Ed.), *Beyond empiricism: Institutions and*

*interventions in the study of crime. Advances in criminological theory.* (pp. 1–37). Transaction. https://doi.org/10.4324/9781351322560-1

Wikström, P.-O. H. (2006). Linking individual, setting and acts of crime. Situational mechanisms and the explanation of crime. In P.-O. H. Wikström & R. J. Sampson (Eds.), *The explanation of crime: context, mechanisms and development* (pp. 61–107). Cambridge University Press.

Wikström, P.-O. H. (2010). Explaining crime as moral actions. In S. Hitlin & S. Vaisey (Eds.), *Handbook of the Sociology of Morality* (pp. 211–239). Springer.

Wikström, P.-O. H. (2014). Why crime happens: A situational action theory. In G. Manzo (Ed.), *Analytical sociology: Actions and networks* (pp. 74–94). Wiley.

Wikström, P.-O. H., & Svensson, R. (2010). When does self-control matter? The interaction between morality and self-control in crime causation. *European Journal of Criminology*, *7*(5), 395–410. https://doi.org/10.1177/1477370810372132

Wikström, P.-O. H., & Treiber, K. (2007). The role of self-control in crime causation: Beyond Gottfredson and Hirschi's general theory of crime. *European Journal of Criminology*, *4*(2), 237–264. ttps://doi.org/10.1177/1477370807074858

Wikström, P.-O. H., & Treiber, K. H. (2009). Violence as situational action. *International Journal of Conflict and Violence*, *3*(1), 75–96. https://doi.org/10.4119/ijcv-2794

Witzel, A., & Reiter, H. (2012). *The problem-centred interview*. SAGE Publications.

Wright, B. R. E., Caspi, A., Moffitt, T. E., & Paternoster, R. (2004). Does the perceived risk of punishment deter criminally prone individuals? Rational choice, self-control, and crime. *Journal of Research in Crime and Delinquency*, *41*(2), 180–213. https://doi.org/10.1177/0022427803260263

Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking.* Cambridge Press.

Zuckerman, M., Eysenck, S. B., & Eysenck, H. J. (1978). Sensation seeking in England and America: cross-cultural, age, and sex comparisons. *Journal of Consulting and Clinical Psychology*, *46*(1), 139–149. https://doi.org/10.1037/0022-006X.46.1.139

## Appendix A: Additional Information on Response Times (RTs)

### Considerations regarding informative value of RTs

Our initial idea was that, in addition to the general and person-specific relevance of the appraisal dimensions, it would also be possible to investigate whether the appraisal dimensions are run through in a fixed sequential order. One possibility would be to conceptualize criminal decision making similar to the emergency intervention process model (Latané & Darley, 1970): The criminal act would stand at the end of a sequence of appraisals; as soon as a certain appraisal turns out negative, the behavioral impulse would be stopped and all subsequent appraisals would be obsolete. Therefore, in our studies we recorded participants' RTs[22] and assumed that -- due to the lower number of appraisals -- RTs would be shorter in case of negative behavioral decisions than in the case of positive ones (i.e., in favor of the criminal action).

In hindsight, we now consider another post-hoc explanation to be more appropriate. Longer RTs in case of a positive (i.e., criminal) response could not be exclusively attributed to the higher number of appraisals, but also to a higher ambivalence of the vignette. If the appraisals on all dimensions are either clearly negative or clearly positive, the decision can be made quickly. However, if some appraisals are negative while others are positive (e.g., high probability of being discovered in combination with a high incentive), the person must ponder how important the respective contradictory appraisals are and, in turn, the decision will take longer. If RTs are on average shorter in decisions against the criminal act, this cannot only be caused by a lower number of appraisals, but also by the fact that those actions were appraised negatively on all dimensions. RTs do not provide a simple readout of the appraisal processes. Therefore, we abandoned this line of thinking and do not discuss RTs any further. Nevertheless, for reasons of transparency, the hypotheses and the corresponding results are presented below.

[22] RTs are defined as the time interval between presentation of the question regarding the behavior execution and the click on one of the answer boxes ("yes" or "no").

**Studies 1 and 2**

*Hypothesis 1: RTs are shorter if participants make negative decisions than if they make positive (i.e., criminal) decisions.*

As preregistered, response times (RTs) that were 1.5 interquartile ratios below the first and above the third quartile of the RT distribution were excluded from the analysis.

A two-sided repeated measures *t*-test revealed a significant difference between RTs in case of a positive decision and RTs in case of a negative decision ($t(295) = -2.64$, $p = .009$): Negative decisions ($M = 2.74$ seconds; $SD = 0.89$) were on average made more quickly than positive ones ($M = 2.90$ seconds; $SD = 1.01$). Hypothesis 1 was therefore confirmed.

*Hypothesis 2: Participants scoring high in psychopathy show shorter RTs in general and especially if they make a positive (i.e., criminal) decision as compared to participants scoring low in psychopathy.*

There was neither a significant correlation (Spearman) between psychopathy (PP-40 total score) and RTs in general ($r(299) = .04$, $p = .46$), nor between psychopathy and RTs in case of a positive response ($r(296) = -.01$, $p = .83$). Thus, the hypothesis was not confirmed.

**Study 3**

*Hypothesis 1: RTs are shorter if participants make negative decisions than if they make positive (i.e., dissexual) decisions.*

The RTs that were 1.5 interquartile ratios below the first and above the third quartile of the RT distribution were excluded from the analysis. A two-sided repeated measures *t*-test revealed a significant difference between RTs in case of a positive decision and RTs in case of a negative decision ($t(648) = -9.22$, $p < .001$): Negative decisions ($M = 3.04$ seconds; $SD = 0.73$) were on average made faster than positive ones ($M = 3.42$ seconds; $SD = 1.08$). Hypothesis 1 was therefore confirmed.

## Appendix B: Preregistrations of Studies 1 and 2

**Preregistration Study 1: https://osf.io/w6c8a**

*Have any data been collected for this study already?*

No, no data have been collected for this study yet

*What's the main question being asked or hypothesis being tested in this study?*

This study tests some of the key assumptions of the appraisal model of criminal decision making. On a PC, participants are presented with small vignettes which offer opportunities for criminal behavior. In a next step, they are asked if they would execute a certain criminal action or not; the answer is given by a binary decision ("yes" or "no") by mouse click. Main hypotheses:

- Response times (RTs) should be shorter when the answer is negative than if the answer is positive
- A manipulation of each of the five dimensions of the appraisal model (impulse strength, legality, morality, likelihood of negative consequences, level of penalty) within the vignettes should lead to a change of the frequency of positive answers.

*Describe the key dependent variable(s) specifying how they will be measured.*

The key dependent variables are RTs measured by the soscisurvey-software, and frequency of positive answers towards potential behavior execution

Additional variables:

- Psychopathy: German short version of the Psychopathic Personality Inventory–Revised (PPI-R-40; Eisenbarth, Lilienfeld & Yarkoni, 2015).
- Big Five personality: German short version of the Big Five Inventory (BFI-K; Rammstedt & John, 2005)
- Demographics and personal information: gender, age, drug consumption during the last 30 days, number of previous convictions, student status

*How many and which conditions will participants be assigned to?*

There are 5 dimensions with 4 pairs of vignettes per dimension, so altogether there are 20 pairs of vignettes (high vs. low value) One pair means there are 2 versions of each vignettes (low vs. high value on the crucial appraisal dimension) Furthermore, there are 20 distractor vignettes (50%) with mundane behavior options which will be presented intermixed with the vignettes for criminal behavior There will be two sets of vignettes to make sure that every participant is presented with either the high value or the low value version of each vignette

*Specify exactly which analyses you will conduct to examine the main question/hypothesis.*

The difference of frequencies will be tested by a chi²-test Difference in RTs between positive and negative answers will be tested by a two-sided t-test Multiple regressions are used to regress RTs on the experimental variable (dummy-coded) and continuous personality variables (psychopathy and Big-Five) for each vignette or parcels of vignettes manipulating the same appraisal dimensions

*Any secondary analyses?*

For each participant the frequency of positive answers will be averaged across both vignettes with a high value on a certain dimension and both with a low value on the same dimension and compared to each other Participants with a high psychopathy score o should show a higher frequency of positive answers for criminal behavior than participants low in psychopathy o should react in a less sensitive way towards a manipulation of the dimension than participants low in psychopathy o should be faster when they make the decision in general or especially when they make a positive decision than participants low in psychopathy Also, it will be explored whether there is a relationship between personality (Big Five), decision making and RTs

*How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.*

We aim at a sample size of $N = 100$ participants

*Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?)*

Parallel to the study at hand, all vignettes will be rated by a group of students regarding their value on each of the five dimensions. The hereby generated normative values will function as a manipulation check regarding the manipulated dimensions Inclusion criterion: Participants must be students or must have studied within the last two years Outliers will be excluded from the analysis: Outliers will be determined using 1,5 IQR below the first and above the 3rd quartile of the RT distribution or if the distribution of RTs shows a clear discontinuity (above which values are considered outliers).

**Preregistration Study 2: https://osf.io/jg6wk**

*Have any data been collected for this study already?*

No, no data have been collected for this study yet

*What's the main question being asked or hypothesis being tested in this study?*

The social-cognitive mechanisms of criminal decision making – AUDIO VERSION. This study tests some of the key assumptions of the appraisal model of criminal decision making. In this online study, participants are presented with AUDIOTAPES of small vignettes which offer opportunities for criminal behavior. In a next step, they are asked if they would execute a certain criminal action or not; the answer is given by a binary decision ("yes" or "no") by mouse click. Main hypotheses:

- Response times (RTs) should be shorter when the answer is negative than if the answer is positive
- A manipulation of each of the five dimensions of the appraisal model (impulse strength, legality, morality, likelihood of negative consequences, level of penalty) within the vignettes should lead to a change of the frequency of positive answers.

*Describe the key dependent variable(s) specifying how they will be measured.*

The key dependent variables are RTs measured by the soscisurvey-software, and frequency of positive answers towards potential behavior execution

Additional variables:

- Psychopathy: German short version of the Psychopathic Personality Inventory–Revised (PPI-R-40; Eisenbarth, Lilienfeld & Yarkoni, 2015).
- Big Five personality: German short version of the Big Five Inventory (BFI-K; Rammstedt & John, 2005)
- Demographics and personal information: gender, age, drug consumption during the last 30 days, number of previous convictions, student status

*How many and which conditions will participants be assigned to?*

There are 5 dimensions with 4 pairs of vignettes per dimension, so altogether there are 20 pairs of vignettes (high vs. low value) One pair means there are 2 versions of each vignettes (low vs. high value on the crucial appraisal dimension) Furthermore, there are 20 distractor vignettes (50%) with mundane behavior options which will be presented intermixed with the vignettes for criminal behavior There will be two sets of vignettes to make sure that every participant is presented with either the high value or the low value version of each vignette

*Specify exactly which analyses you will conduct to examine the main question/hypothesis.*

The difference of frequencies by each appraisal dimension will be tested by a phi-coefficient-test Difference in RTs between positive and negative answers will be tested by a two-sided *t*-test. Multiple regressions are used to regress RTs and positive/negative answers on the experimental variable (dummy-coded) and continuous personality variables (psychopathy and Big-Five) across all vignettes or parcels of vignettes manipulating the same appraisal dimensions.

*Any secondary analyses?*

For each participant, the frequency of positive answers will be averaged across both vignettes with a high value on a certain dimension and both with a low value on the same dimension and compared to each other. Participants with a high psychopathy score

- should show a higher frequency of positive answers for criminal behavior than participants low in psychopathy
- should react in a less sensitive way towards a manipulation of the dimension than participants low in psychopathy
- should be faster when they make the decision in general or especially when they make a positive decision than participants low in psychopathy

Also, it will be explored whether there is a relationship between personality (Big Five), decision making and RTs

*How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.*

We aim at a sample size of $N = 200$ participants

*Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?)*

In a previous study, all vignettes were rated by a group of students regarding their value on each of the five dimensions. The hereby generated normative values will function as a manipulation check regarding the manipulated dimensions. Outliers will be excluded from the analysis: Outliers will be determined using 1,5 IQR below the first and above the 3rd quartile of the RT distribution or if the distribution of RTs shows a clear discontinuity (above which values are considered outliers). Further exclusion criteria: participants who failed the seriousness check (Aust, Ullrich, & Musch, 2013); participants who are neither students nor have studied within the last two years.

## Appendix C: Manipulation Checks Studies 1-3

### Table C1

*Manipulation check results for vignettes used in Studies 1 and 2 (repeated measures ANOVA, within-participant contrasts).*

| **Vignette content** | **Incentive** | | | **Legality** | | | **Morality** | | | **Likelihood of Punishment** | | | **Level of Punishment** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ***F*** | ***p*** | **$\eta^2_P$** | ***F*** | ***p*** | **$\eta^2_P$** | ***F*** | ***p*** | **$\eta^2_P$** | ***F*** | ***p*** | **$\eta^2_P$** | ***F*** | ***p*** | **$\eta^2_P$** |
| Insurance fraud [a] | 1.00 | .36 | .14 | 1.00 | .36 | .14 | 1.00 | .36 | .14 | 4.50 | .08 | .43 | 1.00 | .36 | .14 |
| Student loan fraud [a] | 1.16 | .31 | .10 | 1.00 | .34 | .09 | 0.38 | .55 | .04 | 1.00 | .34 | .09 | 1.00 | .34 | .09 |
| Literature download | 9.59 | **.01**** | .52 | 1.00 | .34 | .10 | 0.00 | 1.00 | .00 | 1.00 | .34 | .10 | 2.25 | .17 | .20 |
| Fake discount | 10.51 | **.01**** | .45 | 0.00 | 1.00 | .00 | 4.50 | .05 | .26 | 1.92 | .19 | .13 | 0.00 | 1.00 | .00 |
| Drunk driving | 0.00 | 1.00 | .00 | 8.71 | **.01**** | .47 | .27 | .61 | .03 | 4.87 | .05 | .33 | 9.41 | **.01**** | .49 |
| Trespassing | 4.50 | .08 | .43 | 9.35 | **.02*** | .61 | 20.25 | **.00**** | .77 | 1.00 | .36 | .14 | 4.50 | .08 | .43 |
| Video download | 5.65 | **.04*** | .39 | 30.41 | **.00***** | .77 | 10.57 | **.01*** | .54 | 15.78 | **.00***** | .64 | 27.92 | **.00***** | .76 |
| Fare dodging train | 10.29 | **.01**** | .46 | 5.80 | **.03*** | .33 | 3.98 | .07 | .25 | 0.00 | 1.00 | .00 | 3.60 | .08 | .23 |
| Faulty bill [a] | 21.78 | **.00***** | .69 | 3.75 | .08 | .27 | 4.87 | .05 | .33 | 0.31 | .59 | .03 | 0.00 | 1.00 | .00 |
| Return used clothes | 8.00 | **.01*** | .38 | 2.17 | .17 | .14 | 10.9 | **.01**** | .46 | 9.75 | **.01**** | .43 | 9.75 | **.01**** | .43 |
| Umbrella theft | 5.33 | **.04*** | .31 | 2.08 | .18 | .15 | 11.64 | **.01**** | .49 | 5.66 | **.04*** | .32 | 9.55 | **.01**** | .44 |
| Alcohol to minors | 9.55 | **.01**** | .44 | 2.51 | .14 | .17 | 9.55 | **.01**** | .44 | 1.93 | .19 | .14 | 8.35 | **.01*** | .41 |
| Traffic light | 4.73 | **.05*** | .25 | 10.42 | **.01**** | .43 | 5.09 | **.04*** | .27 | 139.71 | **.00***** | .91 | 6.14 | **.03*** | .31 |
| Embezzlement | 1.32 | .27 | .09 | - [b] | - | - | 1.00 | .34 | .07 | 40.44 | **.00***** | .76 | 6.30 | **.03*** | .33 |
| Shoplift [a] | 0.00 | 1.00 | .00 | 1.00 | .33 | .06 | 2.13 | .16 | .12 | 1.89 | .19 | .11 | 1.00 | .33 | .06 |
| Waste disposal | 6.00 | **.03*** | .30 | 0.00 | 1.00 | .00 | 1.00 | .33 | .07 | 30.17 | **.00***** | .68 | 9.95 | **.01**** | .42 |
| Speeding | 0.10 | .75 | .01 | 1.00 | .34 | .07 | 2.93 | .11 | .18 | 4.45 | .06 | .26 | 32.5 | **.00***** | .71 |
| Fare dodging tram | 11.42 | **.01**** | .47 | 1.00 | .34 | .07 | 4.50 | .05 | .26 | 2.52 | .14 | .16 | 22.65 | **.00***** | .64 |
| Overtaking | 1.75 | .21 | .11 | 2.15 | .16 | .13 | 3.50 | .08 | .20 | 0.00 | 1.00 | .00 | 26.05 | **.00***** | .65 |
| Property damage [a] | 0.38 | .55 | .03 | 1.00 | .33 | .07 | 0.32 | .58 | .02 | 1.31 | .27 | .09 | 0.32 | .58 | .02 |

*Note:* The values that were to be manipulated are marked by a frame; [a] Vignette was excluded from the analysis of the main study due to failed manipulation check; [b] no values available, as sum of squares was 0; *df*s for all cells = 1; $N$=17; * $p < .05$, ** $p < .01$, *** $p < .001$

**Table C2**

*Manipulation check results for vignettes used in Study 3 (repeated measures ANOVA, within-participant contrasts).*

| **Vignette content** | **Incentive** | | | **Legality** | | | **Morality** | | | **Likelihood of Punishment** | | | **Level of Punishment** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ***F*** | ***p*** | **$\eta^2_P$** | ***F*** | ***p*** | **$\eta^2_P$** | ***F*** | ***p*** | **$\eta^2_P$** | ***F*** | ***p*** | **$\eta^2_P$** | ***F*** | ***p*** | **$\eta^2_P$** |
| Buy drinks to make s.o. compliant | 6.54 | **.02*** | .28 | 1.66 | .22 | .09 | 2.03 | .17 | .11 | 3.77 | .07 | .18 | 2.46 | .14 | .13 |
| Whistling after s.o. | 5.59 | **.03*** | .25 | 0.27 | .61 | .02 | 1.51 | .24 | .08 | 1.35 | .26 | .07 | 3.24 | .09 | .16 |
| Peeking into a locker room | 38.44 | **.00***** | .64 | 1.00 | .33 | .04 | 0.66 | .43 | .03 | 3.15 | .09 | .13 | 1.35 | .26 | .06 |
| Bondage w/o consent [a] | 3.32 | .09 | .16 | 3.77 | .07 | .18 | 3.40 | .08 | .17 | 0.00 | 1.00 | .00 | 2.13 | .16 | .11 |
| Observing neighbor naked | 9.43 | **.01**** | .31 | 10.04 | **.01**** | .32 | 10.93 | **.00**** | .34 | 0.24 | .63 | .01 | 6.83 | **.02**** | .25 |
| Not telling about STD before sex | 7.00 | **.02*** | .26 | 8.87 | **.01**** | .31 | 5.09 | **.04*** | .20 | 5.07 | **.04*** | .20 | 8.44 | **.01**** | .30 |
| Anal penetration w/o consent | 0.48 | .50 | .04 | 20.44 | **.00***** | .63 | 1.57 | .24 | .12 | 5.57 | **.04*** | .32 | 14.49 | **.00**** | .55 |
| Watching rape porn [a] | 14.82 | **.00**** | .51 | 4.37 | .06 | .24 | 6.00 | **.03*** | .30 | 2.15 | .16 | .13 | 3.65 | .08 | .21 |
| Continuing sex w/o consent | 5.09 | **.04*** | .27 | 0.80 | .39 | .05 | 7.98 | **.01*** | .36 | 3.50 | .08 | .20 | 0.27 | .61 | .02 |
| Stealing underwear [a] | 0.39 | .54 | .02 | 0.25 | .62 | .01 | 3.65 | .07 | .15 | 1.51 | .23 | .07 | 1.15 | .30 | .05 |
| Sending penis photo w/o consent | 10.95 | **.01**** | .42 | 1.58 | .23 | .10 | 9.83 | **.01**** | .40 | 0.52 | .48 | .03 | 0.03 | .88 | .00 |
| Watching violent livestream | 12.67 | **.00**** | .40 | 0.00 | 1.00 | .00 | 4.39 | **.05*** | .19 | 4.52 | **.05*** | .19 | 0.00 | 1.00 | .00 |
| Touching breasts w/o consent | 7.00 | **.02*** | .26 | 1.30 | .27 | .06 | 5.40 | **.03*** | .21 | 10.12 | **.01**** | .34 | 0.19 | .67 | .01 |
| Taking nude photos w/o consent | 3.04 | .10 | .13 | 3.32 | .08 | .14 | 1.65 | .21 | .07 | 7.88 | **.01*** | .27 | 2.90 | .10 | .12 |
| Deep throat w/o consent | 1.00 | .33 | .06 | 0.68 | .42 | .04 | 1.52 | .24 | .09 | 4.56 | **.05*** | .22 | 2.71 | .12 | .15 |
| Touching buttocks w/o consent | 1.20 | .29 | .05 | 1.30 | .27 | .06 | 1.87 | .19 | .08 | 29.53 | **.00***** | .58 | 0.04 | .85 | .00 |
| Pressing crotch to s.o.'s buttocks | 0.81 | .38 | .05 | 4.50 | **.05*** | .21 | 0.65 | .43 | .04 | 6.36 | **.02*** | .27 | 31.23 | **.00***** | .65 |
| Filming sex w/o consent | 0.05 | .83 | .00 | 0.04 | .85 | .00 | 0.16 | .69 | .01 | 0.05 | .83 | .00 | 9.59 | **.01**** | .31 |
| Sending sexually suggestive messages | .49 | .49 | .02 | 8.44 | **.01**** | .30 | 0.52 | .48 | .03 | 2.11 | .16 | .10 | 5.05 | **.04*** | .20 |
| Sexist remark to colleague | 0.00 | 1.00 | .00 | 2.03 | .17 | .10 | 1.51 | .24 | .08 | 0.00 | 1.00 | .00 | 10.5 | **.00**** | .37 |

*Note:* The values that were to be manipulated are marked by a frame; [a] Vignette was excluded from the analysis of the main study due to failed manipulation check; *df*s for all cells = 1; *N*=23; * $p < .05$, ** $p < .01$, *** $p < .001$

## Appendix D: Preregistration of Study 3

**Preregistration link Study 3: https://osf.io/n783y**

*Have any data been collected for this study already?*

No, no data have been collected for this study yet.

*What's the main question being asked or hypothesis being tested in this study?*

This study tests some of the key assumptions of the appraisal model of criminal decision making, focusing on the decision-making process in the field of sexual offenses. In this online study, participants are presented with short text-based vignettes which offer opportunities for sexual assault/dissexual behavior. In a next step, they are asked if they would execute the described behavior or not; the answer is given by a binary decision ("yes" or "no") by mouse click.

Main hypotheses:

- Response times (RTs) should be shorter when the answer is negative than if the answer is positive.
- A manipulation of each of the five dimensions of the appraisal model (impulse strength, legality, morality, likelihood of negative consequences, level of costs) within the vignettes should lead to a change of the frequency of positive answers.
- The individual positions of the vignettes should predict the frequency of positive answers due to an assumed rising level of sexual arousal during the task.

*Describe the key dependent variable(s) specifying how they will be measured.*

The key dependent variables are RTs measured by the soscisurvey-software, and frequency of positive answers towards potential behavior execution.

Additional variables:

- Sexual arousal: we assume that participants will become sexually aroused during the task. To test this assumption participants will rate their sexual arousal on a five-point Likert-scale after completing the vignettes. We will also check for non-sexual arousal

and the participants' moods (positive/negative), measured by two self-generated items each.

- Morality: measured by the modified version of Hirtenlehner and Kunz (2016), whereby the response scale was modified to a five-point Likert-scale.
- Self-control: German short version of the Self-Control Scale (SCS-K-D; Bertrams & Dickhäuser, 2009).
- Sex drive: total sexual outlets (Kinsey, Pomeroy, & Martin, 1948), i.e. number of orgasms per week in the last 6 months and maximum number of orgasms per week since participants' 15th birthdays
- Sociosexual orientation: German version of the Revised Sociosexual Orientation Inventory (SOI-R; Penke & Asendorpf, 2008)
- Sexual disgust: subscale of the German 5-Factor Disgust Scale (5-FES, Fünf-Faktoren Ekelskala; Eickmeier, Hoffmann, & Banse, 2017)
- Demographics and personal information: gender, age, education level, sexual orientation (Kinsey-scale, two-item-version; Kinsey, Pomeroy, & Martin, 1948)

*How many and which conditions will participants be assigned to?*

- There are five dimensions with four pairs of vignettes per manipulated dimension, so altogether there are 20 pairs of vignettes (high vs. low value).
- One pair means there are two versions of each vignette (low vs. high value on the crucial appraisal dimension).
- Furthermore, there are 20 distractor vignettes (50%) with mundane behavior options or non-sexual delinquent behavior options which will be presented intermixed with the vignettes for dissexual behavior.

There will be two sets of vignettes to make sure that every participant is presented with either the high value or the low value version of each vignette.

*Specify exactly which analyses you will conduct to examine the main question/hypothesis.*

- Difference in RTs between positive and negative answers will be tested by a t-test.
- The difference of frequencies by each appraisal dimension will be tested by a phi coefficient-test.

- Multiple regressions are used to regress RTs and positive/negative answers on the identity of participants (dummy-coded), the experimental variable (effect-coded), the position of every vignette in the protocol (as a proxy for the situational induction of sexual arousal) and continuous personality variables (sexual arousal, sex-drive, sociosexual orientation, sexual disgust, moral, self-control) across all vignettes or parcels of vignettes manipulating the same appraisal dimensions.

*Any secondary analyses?*

For each participant, the frequency of positive answers will be averaged across both vignettes with a high value on a certain dimension and both with a low value on the same dimension and compared to each other.

- Participants with a high morality score and a high self-control score should show a lower frequency of positive answers for dissexual behavior than participants low in morality/low in self-control.
- Due to a low base rate of endorsing dissexual behavior, participants with a high morality score should show no or only weak responses to a manipulation of the three dimensions impulse strength, likelihood of negative consequences and level of costs in comparison to participants high in morality.
- Participants with a high self-control score should show no or only weak responses to a manipulation of the impulse strength dimension in comparison to participants high in self-contol.

Also, it will be explored whether there is a relationship between sexuality-related variables (sex drive, sociosexual orientation, sexual disgust), decision making and RTs. In particular:

- Participants with a high sexual disgust score/ a high sex drive should show a lower frequency of positive answers for dissexual behavior than participants low in sexual disgust/with a low sex drive.

*How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.*

We aim at a sample size of at least 150 participants. If the data collection is going well, we will exceed this number. However, we will stop collecting data on June 30th, 2018.

*Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?)*

Parallel to the study at hand, all vignettes will be rated by 20 participants regarding their value on each of the five appraisal dimensions. The hereby generated normative values will function as a manipulation check regarding the manipulated dimensions.

Outliers will be excluded from the analysis: Outliers will be determined using 1,5 IQR below the first and above the 3rd quartile of the RT distribution or if the distribution of RTs shows a clear discontinuity (above or below which values are considered outliers).

Further exclusion criteria: participants who failed the seriousness check (Aust, Ullrich, & Musch, 2013); participants who read the vignettes in less than five seconds; female participants; participants who are not sexually interested in women.

## Appendix E: Preregistration of Study 4

**Preregistration link Study 4: https://osf.io/9abf8**

*Have any data been collected for this study already?*

No, no data have been collected for this study yet

*What's the main question being asked or hypothesis being tested in this study?*

We are going to conduct interviews in order to assess whether the dimensions proposed in the appraisal model of criminal decision making (incentive, feasibility, legality, morality, likelihood of negative consequences, and level of penalty) also apply within a sample of offenders and whether the model assumptions can be extended to more serious crime. In addition, we would like to explore if there are additional aspects (appraisal dimensions) which play a role in making criminal decisions in offenders. Additionally, our vignette paradigm, which has already been tested with students, will be used to systematically examine the effects of the appraisal dimensions on fictional criminal behavior. We assume that a manipulation of the dimensions in the vignettes leads to a change in the frequency of the execution of criminal behavior (lower value - more negative answers).

*Describe the key dependent variable(s) specifying how they will be measured.*

- Qualitative content analysis according to Mayring (2008): The categories used correspond to the categories of the appraisal model.
- RTs
- Frequency of positive answers towards potential behavior execution

Additional variables measured:

- Demographics and personal information: gender, age, marital status.
- History of offenses: number of previous convictions, types of offenses which caused previous convictions, type of offense which caused most recent conviction, time spent in prison
- Drug consumption during the last 30 days
- Seriousness check (Aust, Ullrich, & Musch, 2013)

*How many and which conditions will participants be assigned to?*

We examine 5 dimensions (all appraisal dimensions of the model except feasibility) with 4 pairs of vignettes per dimension, so altogether there are 20 pairs of vignettes (high vs. low value) One pair means there are 2 versions of each vignettes (low vs. high value on one specific appraisal dimension) Furthermore, there are 20 distractor vignettes (50%) with mundane behavior options which will be presented intermixed with the vignettes for criminal behavior There will be two sets of vignettes to make sure that every participant is presented with either the high value or the low value version of each vignette

*Specify exactly which analyses you will conduct to examine the main question/hypothesis.*

Qualitative content analysis according to Mayring. The aim is to determine the type of appraisal dimensions preceding criminal/non-criminal behavior in offenders. There will be two coders, so the inter-coder-reliability will be calculated. The frequencies of positive answers will be compared across all vignettes and all dimensions, but also separately for each dimension (for each participant the frequency of positive answers will be averaged across both vignettes with a high value on a certain dimension and both with a low value on the same dimension)

*Any secondary analyses?*

We are going to explore whether the RTs will be influenced by the consistency of the pieces of information presented in the vignettes. The more consistent the information is in regard to the execution or non-execution of the criminal behavior (i.e. the more information favors or contradicts the behavior), the shorter the RTs.

*How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.*

In order to recruit participants, we are cooperating with several institutions, which are linked with offenders (e.g. prison, assisted living, advice center). Our final number of participants will depend on the commitment and the capacities of these institutions. 20-40 participants are expected.

*Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?)*

Outliers will be excluded from the analysis: Outliers in RTs will be determined using 1,5 IQR below the first and above the third quartile of the RT distribution or if the distribution of RTs shows a clear discontinuity (above or below which values are considered outliers). Further exclusion criteria: participants who have failed the seriousness check (Aust, Ullrich, & Musch, 2013); participants who have read the vignettes in less than five seconds. We have already conducted two pilot interviews. It is planned to include these two interviews in the qualitative content analysis. No other data has been collected yet.

## Appendix F: Interview Guide Study 4

**Instruktion:**

*Im Rahmen unseres Forschungsprojektes an der Universität Bonn interessieren wir uns dafür, was in den Köpfen von Personen vorgeht, wenn sie sich für oder gegen eine bestimmte Handlung entscheiden. Das Interview wird ca. 30 Minuten dauern. Wenn Sie das Interview abbrechen möchten, können Sie mir jederzeit Bescheid geben. Die Teilnahme ist selbstverständlich freiwillig. Alles, was wir im Rahmen dieses Interviews besprechen, wird anonym behandelt, also nicht mit Ihrer Person in Verbindung gebracht.*

*Hinweis bzgl. Tonaufnahme: Um Ihre Gedanken nach dem Interview zu erinnern, werde ich das Gespräch aufnehmen. Ich werde das Gespräch auswerten und die Tonaufnahme danach löschen. Die Ergebnisse werden ausschließlich zu Forschungszwecken verwendet. Haben Sie an dieser Stelle noch weitere Fragen?*

**Teil 1:**

*Im Folgenden werde ich Ihnen einige Szenarien vorstellen und Sie anschließend fragen, wie Sie sich in der spezifischen Situation verhalten würden. Sprechen Sie bitte alles aus, was Ihnen in den Sinn kommt und durch den Kopf geht, während Sie sich für oder gegen eine Verhaltensoption entscheiden. Dabei ist es wichtig, dass Sie nicht versuchen Ihre Gedanken zu rechtfertigen. Denken Sie einfach laut. Es gibt keine richtigen und falschen Antworten.*

Sie finden einen 20€-Schein auf der Straße und es ist niemand zu sehen, dem er gehören könnte. Behalten Sie das Geld? (Abwarten, welche Gedanken Pb äußert)

- Falls nein: Unter welchen Bedingungen würden Sie sich dafür entscheiden, das Geld zu behalten? Welche Fragen müssten Sie vorher für sich selbst klären?
- Falls ja: Welcher Gedanke hat Sie schließlich dazu bewogen, es zu tun?

Sie finden einen 200€-Schein auf der Straße und es ist niemand zu sehen, dem er gehören könnte. Behalten Sie das Geld? (Abwarten, welche Gedanken Pb äußert)

- Falls Nein: Unter welchen Bedingungen würden Sie sich dafür entscheiden, das Geld zu behalten? Welche Fragen müssten Sie vorher für sich selbst klären?
- Falls ja: Welcher Gedanke hat Sie schließlich dazu bewogen, es zu tun?

Sie haben legal auf einer von Ihnen abonnierten Streaming-Plattform drei Staffeln Ihrer Lieblingsserie geschaut, für die vierte Staffel hat diese aber keine Lizenz bekommen. Ein Freund hat Ihnen von einer ausländischen Seite erzählt, auf der die 4. Staffel der Serie verfügbar ist. Nutzen Sie den Stream? (Abwarten, welche Gedanken Pb äußert)

- Falls nein: Unter welchen Bedingungen würden Sie sich dafür entscheiden, den Stream zu nutzen? Welche Fragen müssten Sie vorher für sich selbst klären?
- Falls ja: Welcher Gedanke hat Sie schließlich dazu bewogen, es zu tun?

Sie sind auf dem Weg zu einem Termin. Als Sie sich in der Straßenbahn ein Ticket kaufen wollen, bemerken Sie, dass Sie ihr Portemonnaie zu Hause vergessen haben. Bleiben Sie trotzdem in der Bahn? (Abwarten, welche Gedanken Pb äußert)

- Falls nein: Unter welchen Bedingungen würden Sie sich dafür entscheiden, den Stream zu nutzen? Welche Fragen müssten Sie vorher für sich selbst klären?
- Falls ja: Welcher Gedanke hat Sie schließlich dazu bewogen, es zu tun?

Ihr Handy ist kaputt und Sie brauchen dringend ein neues. Vor kurzer Zeit hat Ihnen ein Freund von einem privaten Händler erzählt, der günstige und hochwertige gebrauchte Handys verkauft. Sie kennen den Händler nicht und haben die Vermutung, dass es gestohlene Handys sein könnten.  Kaufen Sie bei dem Händler? (Abwarten, welche Gedanken Pb äußert)

- Falls nein: Unter welchen Bedingungen würden Sie sich dafür entscheiden, den Stream zu nutzen? Welche Fragen müssten Sie vorher für sich selbst klären?
- Falls ja: Welcher Gedanke hat Sie schließlich dazu bewogen, es zu tun?

**Teil 2:**

*Nun würde ich gerne ein Delikt mit Ihnen besprechen, das Sie in der Vergangenheit begangen haben. Wenn es mehrere Delikte gibt, die in Frage kommen, können Sie sich gerne selbst aussuchen, über welches Sie sprechen möchten. Es sollte ein Delikt sein, an das Sie sich noch gut erinnern können.*

*Bitte schildern Sie ausführlich die Situation. Wie ist es zu der Handlung gekommen? Sie können alles beschreiben, was Ihnen einfällt. Es gibt kein Richtig oder Falsch.*

(Teilnehmer zum freien Bericht anregen, aufrechterhaltende Fragen stellen, z.B.:)

- Wie ging es dann weiter?
- Was ist dann passiert?
- Können Sie das noch etwas genauer erklären?
- Was kann ich mir darunter vorstellen? / Wie meinen Sie das genau?
- Können Sie mir dazu Beispiele nennen?

Fragen zur Situation:

- Was ging Ihnen durch den Kopf?
- Welche Gedanken haben Sie sich vor der Handlung gemacht?
- Welche Gedanken haben Sie sich währenddessen gemacht?
- Gab es einen Aspekt, der besonders relevant für die Entscheidung war?
- Haben Sie bei der Entscheidung stärker an angenehme oder unangenehme Dinge gedacht?
- Sind bestimmte Emotionen oder Gefühle aufgetaucht?
- Gab es körperliche Reaktionen, an die Sie sich erinnern?
- Hatten Sie in der Situation eher ein gutes oder schlechtes Gefühl?

Spezifische Fragen zu den Dimensionen:

Anreiz:

- Wie verlockend war die illegale Verhaltensoption / die Ausführung der Straftat für Sie?
- Was hat Sie motiviert?
- Welche Ziele haben beim Entschluss zur Straftat eine Rolle gespielt?
- Gab es eine Sache, die Sie besonders an der Ausführung der Tat gereizt hat?

Machbarkeit:

- Welche Rolle hat die Umsetzbarkeit oder Machbarkeit gespielt? (Ggf. auf Delikt beziehen und Beispiel nennen, wie die Machbarkeit in der Situation eingeschränkt gewesen sein könnte)

Legalität:

- Welche Rolle hatte es für Ihre Entscheidung, dass das Verhalten in der Situation illegal war?

- Gab es legale Alternativen?
- Haben Sie es eher als Hindernis oder Anreiz erlebt, dass die Handlung illegal war?

Moralität:

- Welche Rolle hat Ihr moralisches Verständnis oder ein schlechtes Gewissen bei Ihrer Entscheidung gespielt?
- Gab es innere Konflikte?
- Unter welchen Bedingungen hätten Sie ein schlechtes Gewissen gehabt? Hätte dieses schlechte Gewissen Ihre Entscheidung beeinflusst?

Wahrscheinlichkeit der Strafe:

- Haben Sie daran gedacht, vielleicht erwischt zu werden?
    - Wenn ja: Hatte dieser Gedanke Einfluss auf Ihre Entscheidung?
    - Wenn nein: Hätten Sie sich anders entschieden, wenn Sie damit gerechnet hätten, gefasst zu werden?
- Haben andere Personen Ihr Verhalten beobachtet / davon gewusst?
    - Wenn Ja: Hat das ihre Entscheidung beeinflusst?
    - Wenn Nein: Hätten Sie sich anders entschieden, wenn andere Personen Ihr Verhalten beobachtet hätten / davon gewusst hätten?

Höhe der Strafe:

- Haben Sie darüber nachgedacht, welche Strafe auf Sie zukommen könnte, falls Sie erwischt werden?
- Haben Sie daran gedacht, dass Ihr Verhalten möglicherweise eine Haftstrafe nach sich ziehen könnte?
- Hätten Sie sich anders entschieden, wenn Sie gewusst hätten, dass ihr Verhalten eine Haftstrafe nach sich zieht?
- Wenn Sie heute die Entscheidung noch einmal treffen müssten: Inwiefern spielt es eine Rolle, dass Sie nun eine Haftstrafe verbüßt haben?
- Inwiefern haben Sie Vor- und Nachteile Ihres Handelns abgewogen?

## Appendix G: Preregistration of Study 5 (Pre-Study and Main Study)

**Preregistration link pre-study: https://osf.io/kh9yj**

*Have any data been collected for this study already?*

No, no data have been collected for this study yet

*What's the main question being asked or hypothesis being tested in this study?*

This is a pre-study. Its aim is to investigate how 84 self-constructed behavioral vignettes which describe opportunities for criminal and non-criminal behavior are rated by participants regarding 6 appraisal dimensions (featured in the Appraisal Model of Criminal Decision Making).

*Describe the key dependent variable(s) specifying how they will be measured.*

We will use one rating item per appraisal dimension to determine the properties of each vignette. All dimensions except legality will be rated on a visual analogue scale (slider) from 0-100. Legality will be measured as a binary categorial variable (legally permitted vs. forbidden).

*How many and which conditions will participants be assigned to?*

To prevent fatigue, each participant will be asked to rate only half of the vignettes (k=42). The set of vignettes will be randomly split into two subsets that will be rated by half of the participants, respectively. Within the subsets, the vignettes will be presented in randomized order.

*Specify exactly which analyses you will conduct to examine the main question/hypothesis.*

We don’t have any hypotheses, as we are just exploring the properties of the vignettes (which will be the crucial independent variable for our main study). We will calculate the means and standard deviations of the values assessed.

*Any secondary analyses?*

No

*How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.*

Our data collection will take place via the Prolific Academic platform. We will start with a number of 40 participants (so each vignette will be rated by 20 participants) and will calculate Cronbach's Alpha (α) for each dimension. We aim at $\alpha \geq .80$ for all dimensions. If α is <.80 for one or more dimensions, we will determine how many more participants are needed to reach $\alpha \geq .80$ and continue data collection accordingly within reasonable economic limits. Our budget is limited to £500 (equivalent to about 90 participants). We will use the following of Prolific Academic's prescreening filters: ages 18-35, balanced gender-ratio (50% men, 50% women), nationality: German, current country of residence: Germany, first language: German.

*Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?)*

Additional variables: We will also assess gender, age, education level, family status, previous convictions, and drug consumption during the last 30 days. Exclusion criteria: For the main study we might exclude some vignettes, which were difficult to rate in the pre-study or whose values were not very suitable for our main study. In the pre-study we will exclude participants, if they show signs of careless responding (e.g. no moving of the slider over several vignettes).

**Preregistration link main study: https://osf.io/qud46/**

*Existing data*

No, no data have been collected for this study yet

*Hypothesis*

In this study we aim to explore how person factors interact with the dimensions of the Appraisal Model of Criminal Decision Making. Participants are presented with short vignettes which describe opportunities for criminal and non-criminal behavior. Participants are asked to indicate the probability of whether they would take the opportunity or not on a scale from 0% to 100%. Based on a previous judgement study, an appraisal profile across the six dimensions "incentive/benefit", "feasibility", "legality", "morality", "likelihood of penalty" and "level of penalty" of every vignette has been determined.

First, we expect that participants' probability of taking the opportunity can be predicted by the six dimensions. Thus, incentive, feasibility, legality, morality, likelihood and level of penalty should be significant predictors for the probability of committing (non-)criminal actions. Direction of effects are assumed as follows:

- A higher incentive should lead to a higher probability of taking the offered action ($\beta_{incentive} > 0$)
- A higher feasibility should lead to a higher probability of taking the offered action ($\beta_{feasibility} > 0$)
- The probability of taking a legal offer should be higher than taking an illegal offer ($\beta_{legality} > 0$)
- A higher immorality should lead to a lower probability of taking the offered action ($\beta_{morality} < 0$)
- A higher likelihood of negative consequences should lead to a lower probability of taking the offered action ($\beta_{P(Penalty)} < 0$)
- A higher level of penalty should lead to a lower probability of taking the offered action ($\beta_{penalty} < 0$).

Second, we assume that participants differ in the degree to which their decision making is influenced by the characteristics of the vignettes. Our theoretical assumption is that due to person factors (such as psychopathy, self-control or sensation-seeking) the appraisal process leads to different individual results (e.g. regarding the assessment of how immoral an action is), which in turn should have an influence on the probability of taking the opportunity, which is described in the vignettes. Thus, differences in decision making between participants should be explained by person factors. We hypothesize that person factors moderate the effect of the vignettes' characteristics on decision making. We assume psychopathy, self-control, sensation-seeking, gender and age as predictors to explain for individual differences in criminal decision making. Effects of these factors are expected as follows:

- Men should have a higher probability of taking the offered action than women ($\beta_0$).
- Younger participants should have a higher probability of taking the offered action than older participants ($\beta_0$).

- Participants higher in psychopathy should be less influenced by the morality dimension than people lower in psychopathy ($\beta_{morality}$).
- People lower in self-control should be more influenced by the incentive of an action than people higher in self-control ($\beta_{incentive}$).
- People higher in sensation-seeking should be less influenced by the likelihood of penalty than people lower in sensation-seeking ($\beta_{P(penalty)}$)..

*Dependent variable*

Our dependent variable is participants' self-assessed probability of taking a criminal or non-criminal action offered in small vignettes describing hypothetical situations. Participants will indicate their probability of taking the offered action on a scale from 0% to 100% We will use the following instruments to measure the person factors mentioned above:

- Psychopathy: German version of the Short Dark Triad questionnaire (Malesza, Ostaszewski, Büchner & Kaczmarek, 2017)
- Self-control: German version of the Brief Self-Control Scale (Bertrams & Dickhäuser, 2009)
- Sensation-seeking: "Need for Stimulation"- subscale from the German version of the Need Inventory of Sensation Seeking (Roth, Hammelstein & Brähler, 2014)
- Demographics and additional variables: gender, age, education level, family status, previous convictions, drug consumption during the last 30 days
- Additional information: Machiavellialism and narcissism (German version of the Short Dark Triad questionnaire (Malesza et al., 2017))

*Conditions*

As we are conducting a non-experimental study there won't be any conditions. Each participant will be presented with the same 80 vignettes and will fill out the personality questionnaires at the end of the study.

*Analyses*

We will use a hierarchical linear regression model to predict participant's decisions in probability of taking the offered action. In this model, decisions for every vignette *v* (level 1) are nested within participants *p* (level 2). In the level 1 model, we predict participant's

probability of taking the offered action in the vignettes by the vignettes' dimensional characteristics. Thus, we use the values for "incentive", "feasibility", "legality", "morality", "likelihood of penalty" and "level of penalty" as predictors. The level 1 model can be formalized as follows:

$$decision_{vp} = \beta_{0p} + \beta_{1p} \cdot incentive_v + \beta_{2p} \cdot feasibility_v + \beta_{3p} \cdot legality_v + \beta_{4p} \cdot morality_v + \beta_{5p} \cdot P(penalty)_v + \beta_{6p} \cdot penalty_v + \varepsilon_{vp}$$

As we expect differences in general decision making between participants as well as differences in the influence of the dimensional characteristics of the vignettes on participants' decision making, we use level 2 information to explain these differences between participants. There are person factors which could be theoretically relevant to account for these differences and can be used as level 2 predictors. As described above we use gender and age as level 2 predictors to explain for individual differences in level 1 intercept $\beta_{0p}$:

$$\beta_{0p} = \gamma_{00} + \gamma_{01} \cdot gender + \gamma_{02} \cdot age + \upsilon_{0p}$$

whereby $\gamma_{00}$ describes the overall mean in decision making over every vignette and every subject and $\upsilon_{0p}$ describes the random non-expainable difference between the overall mean and the participant's individual value.

Person factors could also account for differences in individual slopes. As described above, we use psychopathy to explain for individual differences in the slope for morality, self-control to explain for individual differences in the slope for incentive and sensation seeking to explain for individual differences in the slope for the likelihood of penalty:

$$\beta_{1p} = \gamma_{10} + \gamma_{11} \cdot self\text{-}control + \upsilon_{1p}$$

$$\beta_{4p} = \gamma_{40} + \gamma_{41} \cdot psychopathy + \upsilon_{4p}$$

$$\beta_{5p} = \gamma_{50} + \gamma_{51} \cdot sensation\text{-}seeking + \upsilon_{5p}$$

All other slopes are treated as random effects without any predictor variables:

$$\beta_{2p} = \gamma_{20} + \upsilon_{2p}$$

$$\beta_{3p} = \gamma_{30} + \upsilon_{3p}$$

$$\beta_{6p} = \gamma_{60} + \upsilon_{6p}$$

Taking these equations together, the overall model can be formalized as follows:

$decision_{vp} = \gamma_{00} + \gamma_{01} \cdot gender + \gamma_{02} \cdot age + \gamma_{10} \cdot Incentive + \gamma_{11} \cdot self\text{-}control \cdot incentive + \gamma_{20} \cdot feasibility + \gamma_{30} \cdot legality + \gamma_{40} \cdot morality + \gamma_{41} \cdot psychopathy \cdot morality + \gamma_{50} \cdot P(penalty) + \gamma_{51} \cdot P(penalty) \cdot sensation\text{-}seeking + \gamma_{60} \cdot penalty + \upsilon_{0p} + \upsilon_{1p} \cdot incentive + \upsilon_{2p} \cdot feasibility + \upsilon_{3p} \cdot legality + \upsilon_{4p} \cdot morality + \upsilon_{5p} \cdot P(penalty) + \upsilon_{6p} \cdot penalty + \varepsilon_{vp}$

We will use RStudio and the R-package "lme4" to fit the model and estimate parameters. *T*-tests are used to test whether slopes differ from zero or not. An R-script with simulated data and the model specification with lme4 is available at https://osf.io/n3s6d/.

*More analyses*

We will fit different model alternatives and compare each of these models with our theoretical model, using information criteria like AIC, BIC and WAIC. We aim to explore the following alternatives:

- Comparing the theoretical multilevel model with the level 1 model without any random effects
- Establishing the model which best fits the data, using cross-validation
- Model alternatives with further interaction effects between legality and likelihood of penalty as well as legality and level of penalty
- Exploring whether any of the measured person factors can explain variation in slopes other than the ones considered in the theoretical model
- Exploring the influence of further person factors like narcissism and machiavellialism.
- Exploring other interactions between person factors and appraisal dimensions than the ones described in our hypotheses.

*Sample size*

Level 1 sample size was set to = 80, as we have 80 vignettes for every participant. To determine level 2 sample size, we simulated the data by our theoretical model, using standardized regression coefficients of for level 1 effects and for level 2 effects. We repeated the simulation 1000 times with varying values for and determined the number of observations which were needed to detect all the specified effects with a probability of min. 0.9. As a result of this we aim at a sample size of. Furthermore, we aim at a balanced gender ratio (50% women, 50% men) and target participants who are aged 18 - 35.

*Other*

In a pre-study, 84 vignettes were rated by participants regarding their value on each of the six dimensions but every participant only rated half of the vignettes. As participants were randomly assigned to one half of the vignettes, 40 vignettes were rated by participants and the other 40 vignettes by participants. Dimensions were rated on a visual analogue scale from 0-100, except legality which was measured as a categorial variable (legally permitted vs. forbidden). One vignette was excluded because of a programming error. As we aimed for a total of 80 vignettes for this study, we calculated the standard deviation within each dimension for every vignette and additionally excluded the three vignettes with the highest sum of standard deviations across all dimensions. In this study, these pre-ratings are used as predictors for decision making. You can find the pre-study's preregistration as well as the analysis and the data from the pre-study here: https://osf.io/n3s6d/.

We will exclude participants if they show response patterns in the personality scales (i.e. consistently responding with the same answer, ignoring reverse-coded items), if they give implausible answers to the demographic items or, if they indicate that they haven't taken part seriously in the seriousness check item. If there are participants younger than 18 or older than 35, we will calculate and report our analyses with and without these participants.

## Appendix H: Pre-Study Rating Results

### Table H1

*Averaged rating values from the pre-study regarding the six appraisal dimensions*

| Vignette # | Incentive | Feasibility | Legality | Morality | Likelihood of Punishment | Level of Punishment |
|---|---|---|---|---|---|---|
| Vignette 01 | 84.53 (28.24) | 97.63 (9.10) | 5.26 (22.94) | 22.11 (27.01) | 25.21 (30.36) | 20.11 (22.43) |
| Vignette 02 | 83.57 (23.49) | 89.81 (22.23) | 38.10 (49.76) | 74.43 (26.83) | 44.62 (34.28) | 53.33 (33.54) |
| Vignette 03 | 75.26 (32.24) | 82.42 (19.81) | 5.26 (22.94) | 35.53 (28.20) | 43.16 (31.34) | 25.89 (18.38) |
| Vignette 04 | 68.52 (26.76) | 91.67 (17.17) | 0.00 (0.00) | 60.71 (26.00) | 52.43 (35.40) | 57.76 (29.46) |
| Vignette 05 | 85.71 (14.08) | 95.43 (8.03) | 4.76 (21.82) | 85.95 (12.87) | 29.24 (28.24) | 35.62 (23.67) |
| Vignette 06 | 70.79 (27.90) | 89.21 (13.85) | 0.00 (0.00) | 42.58 (32.37) | 43.21 (26.34) | 56.68 (30.10) |
| Vignette 07 | 87.24 (17.28) | 61.71 (32.82) | 0.00 (0.00) | 61.71 (26.57) | 71.38 (26.22) | 67.00 (22.43) |
| Vignette 08 | 70.29 (25.55) | 83.38 (18.55) | 9.52 (30.08) | 45.90 (27.43) | 32.76 (27.36) | 45.14 (26.32) |
| Vignette 09 | 81.90 (27.40) | 97.24 (7.68) | 23.81 (43.64) | 40.90 (27.45) | 21.10 (24.93) | 44.76 (27.80) |
| Vignette 10 | 84.11 (20.29) | 93.05 (11.73) | 63.16 (49.56) | 60.26 (33.45) | 15.68 (20.06) | 19.63 (24.49) |
| Vignette 11 | 72.86 (21.99) | 81.76 (15.41) | 0.00 (0.00) | 86.95 (15.55) | 68.05 (30.57) | 56.14 (22.93) |
| Vignette 12 | 85.84 (21.91) | 95.58 (10.11) | 5.26 (22.94) | 54.63 (32.90) | 19.79 (18.22) | 17.11 (19.13) |
| Vignette 13 | 87.52 (19.64) | 94.24 (11.93) | 0.00 (0.00) | 44.86 (29.30) | 30.14 (27.60) | 42.00 (23.87) |
| Vignette 14 | 84.48 (24.37) | 76.24 (23.52) | 52.38 (51.18) | 62.05 (25.28) | 22.81 (25.26) | 12.14 (16.41) |
| Vignette 15 | 71.71 (28.12) | 73.95 (29.00) | 0.00 (0.00) | 92.48 (9.49) | 56.71 (35.18) | 63.33 (18.71) |
| Vignette 16 | 62.32 (36.71) | 95.47 (11.03) | 5.26 (22.94) | 74.26 (31.35) | 25.47 (22.63) | 32.63 (28.76) |
| Vignette 17 | 76.47 (26.00) | 81.74 (25.23) | 5.26 (22.94) | 62.37 (32.70) | 48.11 (29.01) | 30.53 (26.51) |
| Vignette 18 | 41.89 (35.52) | 92.11 (12.60) | 0.00 (0.00) | 83.53 (19.03) | 59.00 (30.24) | 68.68 (30.92) |
| Vignette 19 | 83.11 (20.59) | 90.74 (17.07) | 0.00 (0.00) | 42.16 (31.45) | 56.47 (25.96) | 43.47 (26.49) |
| Vignette 20 | 55.38 (27.83) | 64.33 (25.29) | 0.00 (0.00) | 48.95 (25.47) | 67.33 (25.70) | 38.52 (21.64) |
| Vignette 21 | 86.14 (13.85) | 91.33 (13.07) | 0.00 (0.00) | 60.62 (25.02) | 42.38 (25.89) | 57.90 (29.78) |
| Vignette 22 | 78.81 (29.22) | 85.71 (21.70) | 0.00 (0.00) | 75.95 (16.17) | 54.52 (31.33) | 59.57 (26.45) |
| Vignette 23 | 92.16 (10.48) | 93.84 (14.60) | 5.26 (22.94) | 23.37 (25.74) | 20.89 (26.50) | 37.37 (29.83) |
| Vignette 24 | 79.81 (20.29) | 55.29 (33.09) | 0.00 (0.00) | 52.29 (30.16) | 69.67 (27.45) | 42.62 (25.82) |
| Vignette 25 | 82.11 (18.28) | 78.47 (19.11) | 21.05 (41.89) | 46.16 (33.05) | 51.11 (26.71) | 48.42 (23.84) |
| Vignette 26 | 19.32 (20.49) | 92.42 (9.70) | 100.00 (0.00) | 71.74 (30.77) | 18.42 (28.77) | 14.11 (20.66) |
| Vignette 27 | 60.24 (33.95) | 88.52 (22.23) | 9.52 (30.08) | 86.95 (14.59) | 46.10 (35.71) | 42.14 (25.38) |
| Vignette 28 | 97.00 (8.94) | 98.52 (5.90) | 100.00 (0.00) | 4.05 (8.92) | 8.10 (8.29) | 1.43 (0.98) |
| Vignette 29 | 85.32 (14.36) | 92.37 (12.20) | 100.00 (0.00) | 51.26 (35.18) | 26.79 (29.78) | 6.53 (9.99) |

**Table H1**

*(continued)*

| Vignette # | Incentive | Feasibility | Legality | Morality | Likelihood of Punishment | Level of Punishment |
|---|---|---|---|---|---|---|
| Vignette 30 | 66.67 (28.38) | 81.24 (12.46) | 4.76 (21.82) | 54.81 (29.71) | 52.57 (31.01) | 31.43 (25.06) |
| Vignette 31 | 21.95 (28.39) | 68.74 (24.73) | 78.95 (41.89) | 49.16 (35.16) | 18.42 (23.02) | 7.63 (12.60) |
| Vignette 32 | 64.79 (34.07) | 86.26 (13.47) | 0.00 (0.00) | 70.16 (23.61) | 44.21 (31.09) | 34.89 (22.39) |
| Vignette 33 | 61.38 (35.53) | 96.48 (8.61) | 100.00 (0.00) | 85.29 (22.94) | 13.81 (17.20) | 10.05 (15.58) |
| Vignette 34 | 75.47 (28.75) | 87.63 (20.19) | 0.00 (0.00) | 89.68 (14.53) | 46.26 (34.40) | 48.63 (28.44) |
| Vignette 35 | 88.52 (13.99) | 78.95 (21.14) | 14.29 (35.86) | 62.00 (30.33) | 38.90 (35.03) | 28.86 (24.86) |
| Vignette 36 | 80.32 (23.62) | 92.53 (17.50) | 5.26 (22.94) | 85.05 (22.31) | 33.37 (29.33) | 36.11 (31.11) |
| Vignette 37 | 71.37 (25.27) | 40.95 (36.05) | 68.42 (47.76) | 76.53 (22.75) | 50.74 (33.99) | 19.74 (23.48) |
| Vignette 38 | 73.52 (26.56) | 99.29 (4.85) | 100.00 (0.00) | 32.57 (23.52) | 20.81 (16.47) | 8.24 (12.02) |
| Vignette 39 | 65.62 (30.16) | 66.86 (20.57) | 0.00 (0.00) | 85.86 (18.86) | 64.10 (30.72) | 71.57 (26.24) |
| Vignette 40 | 35.33 (25.11) | 89.90 (14.44) | 100.00 (0.00) | 62.90 (26.78) | 38.67 (32.33) | 30.00 (30.99) |
| Vignette 41 | 75.14 (24.50) | 83.00 (24.73) | 0.00 (0.00) | 94.57 (8.11) | 60.14 (29.61) | 55.29 (29.22) |
| Vignette 42 | 23.05 (25.19) | 96.63 (10.24) | 100.00 (0.00) | 66.21 (34.00) | 45.42 (34.65) | 6.63 (15.55) |
| Vignette 43 | 86.57 (19.47) | 85.81 (24.52) | 0.00 (0.00) | 82.00 (25.76) | 77.24 (25.47) | 48.52 (27.28) |
| Vignette 44 | 86.11 (24.80) | 92.89 (14.43) | 21.05 (41.89) | 49.63 (32.96) | 39.47 (32.49) | 24.21 (24.33) |
| Vignette 45 | 89.29 (17.57) | 99.05 (6.43) | 100.00 (0.00) | 42.29 (32.98) | 2.86 (5.09) | 1.29 (0.78) |
| Vignette 46 | 90.95 (11.06) | 78.86 (23.40) | 100.00 (0.00) | 38.43 (29.54) | 7.24 (9.90) | 7.52 (15.48) |
| Vignette 47 | 68.53 (33.99) | 89.74 (13.25) | 84.21 (37.46) | 46.00 (34.70) | 32.53 (30.61) | 10.53 (19.16) |
| Vignette 48 | 67.84 (35.13) | 79.89 (22.56) | 31.58 (47.76) | 11.53 (11.63) | 21.84 (19.40) | 17.42 (18.08) |
| Vignette 49 | 70.68 (24.98) | 87.16 (13.27) | 15.79 (37.46) | 77.21 (22.59) | 37.79 (31.77) | 21.63 (26.45) |
| Vignette 50 | 74.71 (26.39) | 24.86 (28.96) | 80.95 (40.24) | 89.29 (11.38) | 58.38 (29.64) | 42.33 (36.73) |
| Vignette 51 | 50.16 (34.55) | 78.21 (30.55) | 100.00 (0.00) | 47.37 (31.36) | 16.11 (16.23) | 8.89 (14.22) |
| Vignette 52 | 71.00 (27.28) | 90.95 (17.77) | 100.00 (0.00) | 24.21 (27.65) | 9.37 (9.39) | 5.00 (9.03) |
| Vignette 53 | 49.67 (29.41) | 80.76 (18.35) | 0.00 (0.00) | 93.29 (11.51) | 57.86 (32.85) | 54.00 (23.04) |
| Vignette 54 | 80.05 (27.33) | 82.16 (22.84) | 26.32 (45.24) | 75.26 (29.09) | 28.79 (32.72) | 31.68 (25.38) |
| Vignette 55 | 72.24 (29.78) | 72.52 (34.55) | 100.00 (0.00) | 57.86 (28.37) | 24.24 (31.60) | 7.38 (10.68) |
| Vignette 56 | 44.16 (27.25) | 75.84 (23.56) | 0.00 (0.00) | 62.32 (30.66) | 63.95 (26.28) | 43.53 (28.18) |
| Vignette 57 | 84.26 (28.64) | 88.89 (25.19) | 0.00 (0.00) | 77.47 (27.86) | 37.53 (30.96) | 42.42 (31.98) |
| Vignette 58 | 55.37 (33.81) | 91.26 (13.76) | 42.11 (50.73) | 74.47 (30.47) | 61.63 (36.41) | 52.84 (30.95) |
| Vignette 59 | 46.29 (28.56) | 80.38 (23.06) | 0.00 (0.00) | 89.71 (13.52) | 71.38 (22.62) | 54.48 (31.34) |
| Vignette 60 | 45.47 (30.87) | 87.11 (23.70) | 26.32 (45.24) | 95.84 (11.23) | 33.00 (35.05) | 43.74 (32.76) |

**Table H1**

*(continued)*

| Vignette # | Incentive | Feasibility | Legality | Morality | Likelihood of Punishment | Level of Punishment |
|---|---|---|---|---|---|---|
| Vignette 61 | 92.62 (9.27) | 83.76 (21.81) | 42.86 (50.71) | 57.90 (28.34) | 55.10 (29.68) | 48.62 (26.38) |
| Vignette 62 | 75.95 (26.11) | 91.24 (13.78) | 0.00 (0.00) | 72.71 (25.40) | 63.71 (31.90) | 82.48 (21.56) |
| Vignette 63 | 68.16 (31.66) | 43.26 (25.95) | 0.00 (0.00) | 90.68 (20.24) | 90.79 (8.12) | 87.74 (13.13) |
| Vignette 64 | 94.29 (8.38) | 94.05 (11.11) | 71.43 (46.29) | 27.95 (26.04) | 22.86 (25.79) | 27.24 (23.62) |
| Vignette 65 | 67.11 (24.12) | 91.11 (14.53) | 0.00 (0.00) | 51.84 (32.58) | 32.11 (26.87) | 24.89 (23.68) |
| Vignette 66 | 78.10 (22.91) | 91.00 (12.39) | 42.86 (50.71) | 73.57 (19.66) | 40.52 (33.64) | 26.71 (23.87) |
| Vignette 67 | 60.29 (32.79) | 96.05 (8.86) | 38.10 (49.76) | 57.33 (29.49) | 14.62 (20.56) | 14.86 (20.39) |
| Vignette 68 | 79.26 (26.34) | 76.16 (28.02) | 15.79 (37.46) | 72.21 (21.61) | 71.58 (26.15) | 35.26 (27.45) |
| Vignette 69 | 67.81 (30.07) | 95.76 (8.68) | 4.76 (21.82) | 40.67 (25.33) | 19.81 (26.29) | 40.76 (32.01) |
| Vignette 70 | 95.48 (8.19) | 97.52 (8.43) | 23.81 (43.64) | 22.29 (24.09) | 11.76 (21.29) | 15.33 (21.45) |
| Vignette 71 | 93.67 (7.40) | 89.43 (16.43) | 0.00 (0.00) | 75.95 (25.32) | 53.24 (33.89) | 60.24 (26.95) |
| Vignette 72 | 40.26 (40.33) | 96.37 (10.37) | 68.42 (47.76) | 55.79 (36.96) | 13.26 (20.05) | 8.79 (8.91) |
| Vignette 73 | 88.33 (15.98) | 78.29 (25.23) | 9.52 (30.08) | 83.62 (17.86) | 69.14 (22.85) | 63.62 (27.09) |
| Vignette 74 | 37.89 (38.66) | 92.68 (13.87) | 21.05 (41.89) | 45.74 (31.55) | 45.16 (30.17) | 46.79 (30.68) |
| Vignette 75 | 94.14 (12.65) | 100.71 (0.64) | 100.00 (0.00) | 3.48 (5.96) | 2.62 (5.19) | 2.81 (5.29) |
| Vignette 76 | 83.53 (26.31) | 69.53 (33.88) | 94.74 (22.94) | 32.21 (33.65) | 13.16 (23.64) | 11.89 (24.72) |
| Vignette 77 | 74.24 (26.94) | 88.38 (21.73) | 100.00 (0.00) | 44.00 (27.11) | 34.10 (26.17) | 10.48 (14.32) |
| Vignette 78 | 76.63 (25.72) | 90.95 (13.13) | 0.00 (0.00) | 66.63 (33.47) | 22.95 (25.51) | 21.68 (24.08) |
| Vignette 79 | 76.84 (24.22) | 91.11 (16.26) | 0.00 (0.00) | 73.11 (25.50) | 34.11 (28.49) | 31.63 (22.57) |
| Vignette 80 | 72.63 (30.32) | 72.00 (26.42) | 15.79 (37.46) | 64.32 (27.78) | 31.63 (23.18) | 30.26 (24.23) |

*Note*. Values in parentheses represent standard deviations. Ratings were made on a continuous scale from 0-100 for all dimensions except *Legality*. *Legality* was measured binary (0 = forbidden, 1 = legally permitted) and rescaled 0-100.

## Appendix I: Model Fit and Evaluation

In total, our final regression model contained 15 fixed effect terms, including three cross-level interactions and eight random effect terms. In order to evaluate the fit of our model, we calculated pseudo-$R^2$ (Nakagawa & Schielzeth, 2013) the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC), and compared these indices between our model and several model alternatives (Table A1). We considered the following model alternatives: 1) the complete model, but in a non-hierarchical regression; 2) the situation model only (Level 1; no person factors) with a random intercept; 3) the complete model (Level 1 and Level 2) with a random intercept; (4) the situation model only (Level 1; no person factors) with random intercept and random slopes; and (5) the complete model with random intercept and random slopes, i.e., our model that was preregistered.

**Table I1**

*Model fits for the preregistered analysis model (i.e., Model #5) and several model alternatives.*

| Model | AIC | BIC | adjusted-/ pseudo-$R^2$ |
|---|---|---|---|
| 1) Complete Model, non-hierarchical | 154,803 | 154,925 | 0.25 |
| 2) Situation Model, random intercept | 154,131 | 154,200 | 0.30 |
| 3) Complete Model, random intercept | 154,075 | 154,205 | 0.30 |
| 4) Situation Model, random slopes | 153,619 | 153,896 | – |
| 5) Complete Model, random slopes (preregistered model) | 153,374 | 153,711 | 0.35 |

*Note.* AIC = Akaike Information Criterion, lower values indicate better fit; BIC = Bayesian Information Criterion, lower values indicate better fit; Calculation of pseudo-$R^2$ for the situation model with random slopes was not possible due to model convergence failures.

The results showed that the non-hierarchical model provided the worst fit to the data, thus, the consideration of the hierarchical structure provided a great benefit. Furthermore, the model alternatives with a random intercept and random slopes provided a better fit than the model alternatives with a random intercept only. Thus, the consideration of differences between participants regarding the influence of the appraisal dimensions (Level 1 predictors) was beneficial. The consideration of person factors (Level 2 predictors) further improved the model fit but only when random slopes were included (i.e., when slopes were allowed to differ between participants). This is in line with the assumption that person factors were accountable for differences in the effect of certain appraisal dimensions on decision making.